Access DB# 119671

SEARCH REQUEST FORM

Scientific and Technical Information Center

69

Requester's Full Name: Luke S Wasson Examiner #: 77895 Date: 19 April 2004
Art Unit: 2177 Phone Number 30 5-5706 Serial Number: 09/992440
Mail Box and Bldg/Room Location: PKZ-4D41 Results Format Preferred (circle): PAPER DISK E-MAIL

If mor than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Method for Generating a Database of Molecular Fragments

Inventors (please provide full names): Richard James Gilbert, William A. Bains, John Caldwell

Earliest Priority Filing Date: 17 November 2000

***For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.**

A method of generating a database of Molecular Fragments, by
starting with a data set of molecular structure data, and iteratively
selecting any two molecular structures from the data set,
comparing the molecular structures;
identifying a molecular fragment that is common to both molecular structures, and
storing the molecular fragment data as a new structure.

Claims also cite the use of graph theory (claim 11), and the creation of 'parent lists' to link molecular fragments to molecular structures which contain them.

R Related to Quantitative Structure-Activity Relationships (QSAR), which through modeling attempts to predict biological characteristics of interest for untested molecules.

Relevant Prior Art attached. Assignee: Amedis Pharmaceuticals LTD

****

STAFF USE ONLY	Type of Search		Vendors and cost where applicable	
Searcher: Holloway	NA Sequence (#)		STN	
Searcher Phone #: 308-7794	AA Sequence (#)		Dialog	# 1589 ✓
Searcher Location: CPK2 4B30	Structure (#)		Questel/Orbit	
Date Searcher Picked Up: 4-19-04	Bibliographic	✓	Dr.Link	
Date Completed: 4-20-04	Litigation		Lexis/Nexis	
Searcher Prep & Review Time: 65	Fulltext	✓	Sequence Systems	✓
Clerical Prep Time:	Patent Family		WWW/Internet	✓
Online Time: 241	Other		Other (specify)	

PTO-1590 (8-01)

STIC Search Report
EIC 2100

STIC Database Tracking Number: 119671

TO: Luke Wassum **Location: 4D41** **Art Unit : 2177** **Tuesday, April 20, 2004** **Case Serial Number: 09/992440**	**From: David Holloway** **Location: EIC 2100** **PK2-4B30** **Phone: 308-7794** **david.holloway@uspto.gov**

Search Notes

Dear Examiner Wassum,

Attached please find your search results for above-referenced case.
Please contact me if you have any questions or would like a re-focused search.

David

QUANTA: Xray Structure and Analysis

B Creating a Fragment Database

Use the following procedure to create a database to be used by the **Search Fragment Database** utility.

1. From the Brookhaven database, select a set of protein coordinates files that have good resolution and include different structure types.

2. Construct a file (dmlist) that contains a list of these protein coordinate files. Use the following format in constructing the file:

- Number of proteins to be used.
- Name of coordinate file 1.
- Name of coordinate file 2.

  .

  .

  .

- Name of coordinate file *n*.

3. Run the program $HYD_MSF/dmprep. The program prompts for the name of the file (dmlist) containing the list of proteins and asks for a name for the distance matrix file (dmfile.new) to be created . The program then reads each protein coordinate file and constructs a distance matrix file. It also creates a QUANTA input command file. The command file is used from within QUANTA to generate an MSF for each of the protein coordinate files. You are prompted to name this file.

The dmprep executable distributed with QUANTA can handle up to 2,000 proteins with limits of 2,000 residues and 100,000 $C\alpha$ distances per protein. The FORTRAN sources for dmprep (dmprep.f and dmsubs.f) are also distributed. This gives you flexibility to increase the dimensions

as you need them.

4. Move the distance matrix file to the $QNT_ROOT/dmatrix directory and rename it to dmfile. Because the variable $HYD_DMF is already defined in the QUANTA environment as $QNT_ROOT/dmatrix/dmfile, you can do this easily by typing:

```
cp dmfile.new $HYD_DMF
```

where dmfile.new is the filename of the distance matrix file created in step 3.

5. To create required MSFs, start QUANTA and type **@command_file**, where command_file is the name given to the QUANTA command file. Respond appropriately to the dialog boxes. Treat the sixth character in the atom field as a disorder using the no-hydrogen dictionary file, and exclude symmetry in the molecular structure file.

6. Move the newly created MSFs to the directory $MSF_LIB.

---

accelrys Previous Next Contents Index Top

Jim's Hom Page

This web page is still under construction.

jdeline@pacbell.net

I am a Ph.D. chemist trained in synthetic organic chemistry. I have been working for The Clorox Company for about the past twelve years, and I live and work in the San Francisco Bay Area of California.

Over the years I have written a couple of software programs for chemists, which I give away for free.

MacFormula
MacFormula is a molecular weight calculator and more. Enter a formula and (optionally) a mass or molar amount, and MacFormula will calculate the molecular weight, % elemental composition, and either a mass or molar amount (depending upon your optional input). Great for planning reactions.

Comes in both a Macintosh and Windows 95 version. The Windows version is called "WinFormula."

Download MacFormula Download WinFormula

MF Calc ("Molecular Fragment Calculator")
MF Calc will take a user defined mass and calculate all of the possible elemental combinations possible with that mass. The program is very flexible in that the user can control the degree of the precision of the mass, as well as which elements (and the amounts) that should be included in the search. Will calculate an exact formula from a high-resolution mass spec value. Available in both Mac and Windows versions.

Download MF Calc (Mac version) Download MF Calc (Windows version)

```
Set     Items   Description
S1       5443   AU=(GILBERT R? OR GILBERT, R?)
S2        210   AU=(BAINS W? OR BAINS, W?)
S3       4387   AU=(CALDWELL J? OR CALDWELL, J?)
S4          7   S1 AND S2 AND S3
S5       1771   (S1 OR S2 OR S3) AND (MOLECUL? OR AMINO? OR GENETIC? OR SE-
             QUENC?)
S6        427   S5 AND (STRUCTUR? OR BOND? OR FRAGMENT? OR PROBE?)
S7          7   S6 AND (DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK?) OR
              DB OR DBMS OR RDB? OR OODB?)
S8          2   S7 AND (QUERY? OR QUERIES OR QUERIED OR LOCAT? OR MATCH? OR
              FIND? OR SEARCH? OR COMPAR?)
S9         14   S4 OR S7
S10         8   RD (unique items)
File   2:INSPEC 1969-2004/Apr W2
         (c) 2004 Institution of Electrical Engineers
File   6:NTIS 1964-2004/Apr W3
         (c) 2004 NTIS, Intl Cpyrght All Rights Res
File   8:Ei Compendex(R) 1970-2004/Apr W2
         (c) 2004 Elsevier Eng.  Info. Inc.
File 148:Gale Group Trade & Industry DB 1976-2004/Apr 19
         (c)2004 The Gale Group
File  94:JICST-EPlus 1985-2004/Apr W1
         (c)2004 Japan Science and Tech Corp(JST)
File 154:MEDLINE(R) 1990-2004/Apr W2
         (c) format only 2004 The Dialog Corp.
File 160:Gale Group PROMT(R) 1972-1989
         (c) 1999 The Gale Group
File  35:Dissertation Abs Online 1861-2004/Mar
         (c) 2004 ProQuest Info&Learning
File  65:Inside Conferences 1993-2004/Apr W2
         (c) 2004 BLDSC all rts. reserv.
File  34:SciSearch(R) Cited Ref Sci 1990-2004/Apr W2
         (c) 2004 Inst for Sci Info
File 315:ChemEng & Biotec Abs 1970-2004/Mar
         (c) 2004 DECHEMA
File 314:CA SEARCH(R) 1997-2004/UD=14017
         (c) 2004 American Chemical Society
File 285:BioBusiness(R) 1985-1998/Aug W1
         (c) 1998 BIOSIS
File  55:Biosis Previews(R) 1993-2004/Apr W2
         (c) 2004 BIOSIS
File  73:EMBASE 1974-2004/Apr W2
         (c) 2004 Elsevier Science B.V.
```

10/5/5 (Item 2 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

09381802 PMID: 1637748

Protein structure prediction from predicted residue properties utilizing a digital encoding algorithm.

Gilbert R J
British Bio-technology Limited, Cowley, Oxford, UK.

Journal of molecular graphics (UNITED STATES) Jun 1992, 10 (2) p112-9, ISSN 0263-7855 Journal Code: 9014762
Document type: Journal Article
Languages: ENGLISH
Main Citation Owner: NLM
Record type: Completed
Subfile: INDEX MEDICUS

Although many disparate methods have been applied to the problem, the accuracy of protein **structural** prediction still remains disappointingly low, averaging about 65% correct secondary **structure** assignment. A novel predictive method is presented here, which attempts to address some of the shortfalls inherent in representing a protein as a simple text-like **sequence** of **amino** acids, by deriving pattern-matching data from the predicted physical properties of a protein chain rather than from the **sequence** itself. A unique binary encoding algorithm is used to enable the property profiles to be correlated with known secondary **structure** , and hence to predict secondary **structures** for proteins with unknown **structures** . By treating the **sequence** in this manner, predictive accuracies averaging over 75% have been achieved.

Descriptors: *Algorithms; *Protein Conformation; **Amino** Acid **Sequence** ; Computer Simulation; **Databases** , Factual; **Molecular Sequence** Data
Record Date Created: 19920903
Record Date Completed: 19920903

```
Set     Items   Description
S1        424   AU=(GILBERT R? OR GILBERT, R?)
S2         25   AU=(BAINS W? OR BAINS, W?)
S3        214   AU=(CALDWELL J? OR CALDWELL, J?)
S4          3   S1 AND S2 AND S3
S5        119   (S1 OR S2 OR S3) AND (MOLECUL? OR AMINO? OR GENETIC? OR SE-
            QUENC?)
S6         76   S5 AND (STRUCTUR? OR BOND? OR FRAGMENT? OR PROBE?)
S7          1   S6 AND IC=G06F-007?
S8          6   S6 AND IC=G06F?
S9         65   S6 AND (MATCH? OR COMPAR? OR QUERY OR QUERIES OR QUERYING -
            OR QUERIED OR RETRIEV? OR ORGANI? OR INDEX? OR INDICE?)
S10        11   S9 AND (DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK? OR
            FILE?) OR DBMS OR RDBMS? OR DB OR OODB? OR DBS)
S11        13   S10 OR S8 OR S7 OR S4
S12        13   IDPAT (sorted in duplicate/non-duplicate order)
S13         9   IDPAT (primary/non-duplicate records only)
File 344:Chinese Patents Abs Aug 1985-2004/Mar
         (c) 2004 European Patent Office
File 347:JAPIO Nov 1976-2003/Dec(Updated 040402)
         (c) 2004 JPO & JAPIO
File 348:EUROPEAN PATENTS 1978-2004/Apr W02
         (c) 2004 European Patent Office
File 349:PCT FULLTEXT 1979-2002/UB=20040415,UT=20040408
         (c) 2004 WIPO/Univentio
File 350:Derwent WPIX 1963-2004/UD,UM &UP=200425
         (c) 2004  Thomson Derwent
```

13/5/2 (Item 2 from file: 350)
DIALOG(R)File 350:Derwent WPIX

014659214 **Image available**
WPI Acc No: 2002-479918/200251
XRAM Acc No: C02-136618
XRPX Acc No: N02-378979

Molecular fragment database **generation method for drugs, involves determining** molecular fragments **that are found within** molecules **of the data set**

Patent Assignee: AMEDIS PHARM LTD (AMED-N)
Inventor: **BAINS W A ; CALDWELL J ; GILBERT R J**

Number of Countries: 099 Number of Patents: 003
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 200241179	A2	20020523	WO 2001GB5096	A	20011116	200251	B
US 20020062307	A1	20020523	US 2001992440	A	20011116	200251	
AU 200215128	A	20020527	AU 200215128	A	20011116	200261	

Priority Applications (No Type Date): GB 200028157 A 20001117
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200241179	A2	E	31	G06F-017/30	

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZM ZW
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW

US 20020062307	A1		G06F-007/00	
AU 200215128	A		G06F-017/30	Based on patent WO 200241179

Abstract (Basic): WO 200241179 A2

NOVELTY - Two **molecular structure** data selected from a data set are **compared** to determine **molecular fragment** data which is then stored. The process is repeated in which one of the **molecular structure** data is selected from either the predetermined **molecular structure** data or the determined **molecular fragment** data, such that the resultant data set is stored in a **database** .

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for:

(1) **molecular fragments** and biological target characteristics relationship determining method;

(2) automated predicted biological target characteristic data generation method; and

(3) predicted biological target characteristic data generation program.

USE - For generating **molecular fragments** relating to drugs.

ADVANTAGE - Since the **molecular fragments** that are actually found within the **molecules** of the data set are determined, time is not wasted in considering entities which are not present. The method is not limited to any particular type of **molecular structure** . The **database** provides the potential for improved data upon which subsequent modeling is performed.

DESCRIPTION OF DRAWING(S) - The figure shows a flow diagram explaining the **molecular fragment database** generation method.

pp; 31 DwgNo 1/4

Title Terms: **MOLECULAR ; FRAGMENT ; DATABASE** ; GENERATE; METHOD; DRUG; DETERMINE; **MOLECULAR ; FRAGMENT** ; FOUND; **MOLECULAR** ; DATA; SET
Derwent Class: B04; T01
International Patent Class (Main): **G06F-007/00 ; G06F-017/30**
File Segment: CPI; EPI

13/5/4 (Item 4 from file: 348)
DIALOG(R)File 348:EUROPEAN PATENTS

01485205
METHOD FOR GENERATING A DATABASE OF MOLECULAR FRAGMENTS
VERFAHREN ZUM ERSTELLEN VON DATENBANKEN FUR MOLEKULARFRAGMENTE
PROCEDE DE GENERATION D'UNE BASE DE DONNEES DE FRAGMENTS MOLECULAIRES
PATENT ASSIGNEE:
Amedis Pharmaceuticals Limited, (4114660), Unit 209, Cambridge Science Park, Milton Road, Cambridge CB4 0GZ, (GB), (Applicant designated States: all)
INVENTOR:
GILBERT, Richard James, Amedis Pharmaceuticals Ltd , 12 St James' Square , London SW1Y 4RB, (GB)
BAINS, William A., Amedis Pharmaceuticals Limited , 12 St James' Square, London SW1Y 4RB, (GB)
CALDWELL, J, Imperial College School of Medicine , Sir Alexander Fleming Bldg, Imperial College Rd, London SW7 2AZ, (GB
PATENT (CC, No, Kind, Date): WO 2002041179 020523
APPLICATION (CC, No, Date): EP 2001983706 011116; WO 2001GB5096 011116
PRIORITY (CC, No, Date): GB 28157 001117
DESIGNATED STATES: AT; BE; CH; CY; DE; DK; ES; FI; FR; GB; GR; IE; IT; LI; LU; MC; NL; PT; SE; TR
EXTENDED DESIGNATED STATES: AL; LT; LV; MK; RO; SI
INTERNATIONAL PATENT CLASS: **G06F-017/30**
LEGAL STATUS (Type, Pub Date, Kind, Text):
Application: 021106 A2 International application. (Art. 158(1))
Application: 021106 A2 International application entering European phase
Application: 040114 A2 International application. (Art. 158(1))
Appl Changed: 040114 A2 International application not entering European phase
Withdrawal: 040114 A2 Date application deemed withdrawn: 20030618
LANGUAGE (Publication,Procedural,Application): English; English; English

```
S1    1154951   MOLECULE? OR MOLECULAR OR PROTEIN? OR PEPTIDE? OR AMINO()A-
             CID? OR GENETIC? OR POLYPEPTIDE?
S2     164262   DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK? OR FILE?) -
             OR DB OR DBS OR DBMS OR RDB OR RDBM OR OODB?
S3    2117122   MATCH? OR COMPAR? OR QUERY OR QERIE? OR QUERYING OR SEARCH?
              OR LOCAT? OR FIND? OR SEEK?
S4    1252118   REPEAT? OR ITERAT? OR REITERAT? OR AGAIN?
S5     484649   GRAPH? OR PARENT? OR INDEX OR INDICE? OR LIST? ?
S6    5904006   FRAGMENT? OR CLIQUE? OR PART OR PARTS OR PARTIAL OR SECTIO-
             N? OR STRING? OR SUBSTRING? OR MF OR MFS OR RESIDUE? OR CHAIN?
S7        454   S1 AND S2 AND S3 AND S4 AND S5 AND S6
S8       2724   S2(2N)(CREAT? OR MAKE? OR DEVELOP? OR POPULAT? OR FILL?)
S9         61   S2(3N)S6(3N)S4
S10         3   S7 AND S8
S11        10   S7 AND S9
S12        20   S1 AND S2 AND S3 AND S4 AND S6 AND S9
S13        13   S1 AND S2 AND S3 AND S4 AND S6 AND S8
S14        33   S10 OR S11 OR S12 OR S13
S15         7   S14 AND IC=G06F?
S16         7   IDPAT (sorted in duplicate/non-duplicate order)
S17         7   IDPAT (primary/non-duplicate records only)
S18        23   S10 OR S11 OR S12
S19        18   S18 NOT S17
S20        18   IDPAT (sorted in duplicate/non-duplicate order)
S21        18   IDPAT (primary/non-duplicate records only)
File 347:JAPIO Nov 1976-2003/Dec(Updated 040402)
        (c) 2004 JPO & JAPIO
File 350:Derwent WPIX 1963-2004/UD,UM &UP=200425
        (c) 2004  Thomson Derwent
```

17/5/1 (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX

015835147
WPI Acc No: 2003-897351/200382
Related WPI Acc No: 2002-435445; 2002-692206; 2003-156963; 2003-201222; 2003-343556; 2003-852784; 2004-010668; 2004-099568
XRAM Acc No: C03-254732
XRPX Acc No: N03-716201

Identification of gene clusters e.g. conferring drug resistance in microorganisms, involves computerized screening of genomic DNA fragments against **known gene cluster** databases **and use of identified** fragments **for cluster detection**

Patent Assignee: ECOPIA BIOSCIENCES INC (ECOP-N); FARNET C M (FARN-I); STAFFA A (STAF-I); ZAZOPOULOS E (ZAZO-I)
Inventor: FARNET C M; STAFFA A; ZAZOPOULOS E
Number of Countries: 098 Number of Patents: 009
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030138810	A1	20030724	US 2000239924	P	20001013	200382 B
			US 2001286346	P	20010426	
			US 2001291959	P	20010521	
			US 2001296744	P	20010611	
			US 2001910813	A	20010724	
			US 2001307629	P	20010726	
			US 2001976059	A	20011015	
			US 2001334604	P	20011203	
			US 2001342133	P	20011226	
			US 2002372789	P	20020417	
			US 2002132134	A	20020426	
			US 2002152886	A	20020521	
			US 2002166087	A	20020611	
			US 2002205032	A	20020726	
			US 2002232370	A	20020903	
CA 2412226	A1	20030622	CA 2412226	A	20021224	200382
CA 2412627	A1	20030626	CA 2412627	A	20021224	200382
WO 200360127	A2	20030724	WO 2002CA2021	A	20021224	200382
WO 200360128	A2	20030724	WO 2002CA2022	A	20021224	200382
CA 2444812	A1	20020904	CA 2387401	A	20020521	200403
			CA 2444812	A	20020521	
CA 2444802	A1	20020904	CA 2387401	A	20020521	200407
			CA 2444802	A	20020521	
AU 2002351637	A1	20030730	AU 2002351637	A	20021224	200421
AU 2002351636	A1	20030730	AU 2002351636	A	20021224	200421

Priority Applications (No Type Date): US 2002232370 A 20020903; US 2000239924 P 20001013; US 2001286346 P 20010426; US 2001291959 P 20010521 ; US 2001296744 P 20010611; US 2001910813 A 20010724; US 2001307629 P 20010726; US 2001976059 A 20011015; US 2001334604 P 20011203; US 2001342133 P 20011226; US 2002372789 P 20020417; US 2002132134 A 20020426 ; US 2002152886 A 20020521; US 2002166087 A 20020611; US 2002205032 A 20020726

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030138810	A1		29	C12Q-001/68	Provisional application US 2000239924
					Provisional application US 2001286346
					Provisional application US 2001291959
					Provisional application US 2001296744
					CIP of application US 2001910813
					Provisional application US 2001307629
					CIP of application US 2001976059
					Provisional application US 2001334604
					Provisional application US 2001342133
					Provisional application US 2002372789
					CIP of application US 2002132134
					CIP of application US 2002152886

CIP of application US 2002166087
CIP of application US 2002205032

CA 2412226 A1 E C12N-015/12
CA 2412627 A1 E C12N-015/12
WO 200360127 A2 E C12N-015/52
Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PH PL PT RO RU SD SE SG SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SI SL SZ TR TZ UG ZW
WO 200360128 A2 E C12N-015/52
Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PH PL PT RO RU SD SE SG SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
Designated States (Regional): AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SI SK SL SZ TR TZ UG ZM ZW
CA 2444812 A1 E C12N-015/55 Div ex application CA 2387401
CA 2444802 A1 E C12N-015/55 Div ex application CA 2387401
AU 2002351637 A1 C12N-015/52 Based on patent WO 200360128
AU 2002351636 A1 C12N-015/52 Based on patent WO 200360127

Abstract (Basic): US 20030138810 A1

NOVELTY - Identifying gene clusters, comprising preparing small- and large-insert libraries of DNA **fragments** from genomic DNA, sequencing **fragments** from the small-insert library, **comparing** using computerized methods to a **database** of known gene clusters to identify **fragments** with similar sequences, and using **fragments** identified to detect gene clusters from the large-insert library, is new.

DETAILED DESCRIPTION - Identifying gene clusters, comprising:

(a) preparing small-insert and large-insert libraries respectively of DNA **fragments** from genomic DNA;

(b) determining DNA sequence of at least **part** of some of the **fragments** in the small-insert library to form Gene Sequence Tags (GSTs);

(c) **comparing** , under computer control, GSTs or corresponding **amino acid** sequences with sequences in a **database** containing genes, gene **fragments** or DNA/ **amino acid** sequences known to be **part** of a gene cluster to identify GSTs with similar structure to a **database** sequence; and

(d) using an identified GST to detect a DNA **fragment** from the large-insert library containing the GST and a gene cluster.

An INDEPENDENT CLAIM is also included for a similar method in which only a large-insert library is prepared and GSTs are identified from the large-insert library.

USE - The method is useful to identify gene clusters associated with a pathogenicity island (i.e. group of genes conferring pathogenicity), degradation of a compound or conferring resistance to a therapeutic drug, especially in cultured/uncultured microorganisms, particularly prokaryotes e.g. of genus Nocardia, Streptomyces, Stigmatella etc. (claimed). It is useful to detect gene clusters involved in biosynthesis of natural products e.g. to identify biosynthetic loci associated with particular products, distinguish between variations of natural products (e.g. between avilamycin-type and everninomycin-type orthomycins) or to identify biosynthetic loci in organisms not previously known to produce the product.

pp; 29 DwgNo 0/5

Title Terms: IDENTIFY; GENE; CLUSTER; CONFER; DRUG; RESISTANCE; MICROORGANISM; COMPUTER; SCREEN; GENOME; DNA; **FRAGMENT** ; GENE; CLUSTER; IDENTIFY; **FRAGMENT** ; CLUSTER; DETECT

Derwent Class: B04; D16; S03; T01

International Patent Class (Main): C12N-015/12; C12N-015/52; C12N-015/55; C12Q-001/68

International Patent Class (Additional): C07H-021/00; C07K-014/195; C07K-014/36; C07K-014/366; C07K-014/47; C07K-016/12; C07K-016/122; C07K-016/40; C07K-016/400; C12N-001/21; C12N-001/211; C12N-009/00;

17/5/2 **(Item 2 from file: 350)**
DIALOG(R)File 350:Derwent WPIX

015584038 **Image available**
WPI Acc No: 2003-646195/200361
XRAM Acc No: C03-176857
XRPX Acc No: N03-514003
Analyzing a biochemical sequence database **by carrying out alignment of a** query **sequence** against **the** database **, and if any result sequences are found and unless a stop condition is met, automatically** repeating **steps using result sequences**
Patent Assignee: DEVGEN NV (DEVG-N)
Inventor: VAN CRIEKINGE W
Number of Countries: 102 Number of Patents: 002
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 200365247	A2	20030807	WO 2003EP1031	A	20030131	200361	B
AU 2003206820	A1	20030902	AU 2003206820	A	20030131	200422	

Priority Applications (No Type Date): US 2002353570 P 20020201; GB 20022398 A 20020201
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200365247	A2	E	34	G06F-017/30	

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SC SD SE SG SK SL TJ TM TN TR TT TZ UA UG US UZ VC VN YU ZA ZM ZW
Designated States (Regional): AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR HU IE IT KE LS LU MC MW MZ NL OA PT SD SE SI SK SL SZ TR TZ UG ZM ZW

AU 2003206820 A1 G06F-017/30 Based on patent WO 200365247

Abstract (Basic): WO 200365247 A2
NOVELTY - Analyzing a biochemical sequence **database** comprises:
(a) providing an initial **query** sequence;
(b) carrying out an alignment of the **query** sequence **against** the **database** to establish result sequences which resemble the **query** sequence according to a measure of similarity; and
(c) if any result sequences are established and unless a stop condition is met, automatically **repeating** the second and third steps using each of the result sequences as a **query** sequence.
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for:
(1) a computer program product comprising computer program instructions to control a computer to carry out the method;
(2) a computer readable medium carrying a computer program product;
(3) an apparatus for analyzing a biochemical sequence **database** , which comprises:
(a) a data store holding the **database** ;
(b) an input arranged to provide an initial **query** sequence;
(c) an alignment engine arranged to carry out an alignment of a **query** sequence **against** the **database** to establish result sequences which resemble the **query** sequence according to a measure of similarity; and
(d) control logic arranged to pass the initial **query** sequence to the alignment engine and to subsequently and **iteratively** pass selected ones of the result sequences to the alignment engine, if any result sequences are established and until a stop condition is met; and
(4) a computer system for carrying out analysis of a biochemical sequence **database** , which comprises:
(a) a storage area network adapted to store the **database** ;
(b) alignment nodes, each operable in response to an instruction to carry out the alignment of a **query** sequence **against** at least a **part** of the **database** ;
(c) a file server connected to the storage area network and to the alignment nodes; and

(d) a head node connected to the file server and to each alignment node and operable to receive initial **query** sequence and to instruct each alignment node to carry out an alignment of the initial **query** sequence **against** the **database** , and operable to receive result sequences from the alignment nodes and to instruct each alignment node to carry out an alignment of a received result sequence **against** the **database** to obtain further result sequences.

USE - The method is used for analyzing a biochemical sequence **database** . It is used e.g., for **comparing** a sequence or set of sequences from a vertebrate organism e.g. fish (e.g., zebrafish), a bird, and/or a mammal (e.g. a mouse, rabbit, rat, monkey, or human) with a **data base** of sequences form an invertebrate organism e.g., an insect (e.g., Drosophila melanogaster) or a nematode, or vice versa.

ADVANTAGE - The provision of at least two different types of subnode in a heterogeneous cluster allows cost and performance to be balanced as required, and reduces the cost of achieving a given level of performance when carrying out a recursive alignment.

DESCRIPTION OF DRAWING(S) - The figure is a flow diagram illustrating a recursive alignment method for analyzing a biochemical sequence **database** .

pp; 34 DwgNo 1/6

Title Terms: BIOCHEMICAL; SEQUENCE; **DATABASE** ; CARRY; ALIGN; **QUERY** ; SEQUENCE; **DATABASE** ; RESULT; SEQUENCE; FOUND; STOP; CONDITION; AUTOMATIC ; **REPEAT** ; STEP; RESULT; SEQUENCE

Derwent Class: B04; D16; S05; T01

International Patent Class (Main): **G06F-017/30**

File Segment: CPI; EPI

17/5/4 (Item 4 from file: 350)
DIALOG(R)File 350:Derwent WPIX

014483718 **Image available**
WPI Acc No: 2002-304421/200234
XRAM Acc No: C02-088615
XRPX Acc No: N02-238158
Computer-readable structure, useful for organizing database **elements corresponding to** proteins **in tissue obtained from organism, comprises records, parameter field,** location **field and abundance field**
Patent Assignee: LARGE SCALE PROTEOMICS CORP (LARG-N); ANDERSON N G (ANDE-I); ANDERSON N L (ANDE-I)
Inventor: ANDERSON N G; ANDERSON N L; ANDERSON N
Number of Countries: 097 Number of Patents: 007
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 200221428	A1	20020314	WO 2001US26933	A	20010831	200234	B
US 20020028005	A1	20020307	US 2000654133	A	20000901	200234	
			US 2001753678	A	20010104		
US 20020087273	A1	20020704	US 2001753678	A	20010104	200247	
			US 2001756285	A	20010109		
AU 200188501	A	20020322	AU 200188501	A	20010831	200251	
US 20030009293	A1	20030109	US 2001756285	A	20010109	200311	
			US 2002235649	A	20020906		
US 20030059095	A1	20030327	US 2000654133	A	20000901	200325	
			US 2002295840	A	20021118		
US 6539102	B1	20030325	US 2000654133	A	20000901	200325	

Priority Applications (No Type Date): US 2001756285 A 20010109; US 2000654133 A 20000901; US 2001753678 A 20010104; US 2002235649 A 20020906 ; US 2002295840 A 20021118
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200221428	A1	E	93	G06K-009/00	

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PH PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

US 20020028005	A1	G06K-009/00	CIP of application US 2000654133
US 20020087273	A1	G06F-019/00	CIP of application US 2001753678
AU 200188501	A	G06K-009/00	Based on patent WO 200221428
US 20030009293	A1	G06F-017/60	Cont of application US 2001756285
US 20030059095	A1	G06K-009/00	Cont of application US 2000654133
US 6539102	B1	G06K-009/00	

Abstract (Basic): WO 200221428 A1
NOVELTY - A computer-readable structure comprising records for storing different types of data relating to respective **proteins** , a parameter field for indicating a selected characteristic of the corresponding **protein** , a **location** field for indicating the relative **location** in the organism from which the **protein** was obtained, and an abundance field for indicating the relative amount of the **protein** , is new.
DETAILED DESCRIPTION - A computer-readable structure, encoded on a computer-readable medium, comprises records for storing different types of data relating to respective **proteins** , a parameter field for indicating a selected characteristic of the corresponding **protein** , a **location** field for indicating the relative **location** in the organism from which the corresponding **protein** was obtained, and an abundance field for indicating the relative amount of the corresponding **protein** obtained from the **location** , where each record has at least an identification field for identifying a corresponding one of the **proteins** , is new.
INDEPENDENT CLAIMS are also included for the following:
(1) a computer program product for extracting selected data

relating to a **protein** from a **database** comprising a computer-readable medium, a user interface module for guiding a user to generate at least one **query** to retrieve selected data from the **database** , a **database search** module communicatively coupled to the user interface module and operable to **locate** and retrieve the **database** that correspond to the **query** ;

(2) determining the proteome of an individual comprising taking a **protein** containing sample from each of at least 5 tissue from an individual and determining the presence and relative abundance of at least 10 **proteins** from each of the tissues;

(3) identifying a **protein** marker that indicates a condition by change in abundance comprising determining the abundance of a candidate **protein** marker in the same biological samples that have different selected characteristic(s), accessing a **database** comprising entries for providing data relating to **proteins** including the candidate **protein** marker, and **comparing** the abundance of the candidate **protein** marker to the entries in the **database** ;

(4) obtaining proteomic information comprising generating a **query** to retrieve selected data relating to a **protein** from the computer program, **locating** a record in the **protein index database** that satisfies **protein** characteristics requested via the **query** and generating an output corresponding to the record;

(5) identifying component-specific **proteins** from a **database** comprising information relating to a number of **proteins** comprising:

(a) generating a first **list** of all **proteins** indicated in the **database** as being **located** in a specimen of a first selected component;

(b) generating a second **list** of all **proteins** indicated in the **database** as being **located** in a specimen of a second selected component;

(c) subtracting from the first **list** all of the **proteins** common to both **lists** ; and

(d) **repeating** steps (b) and (c) for components 3-n, where n is the total number of components in the **database** 6) **creating** a **polypeptide database** comprising:

(a) generating a 2-D separation of **polypeptides** of two sources;

(b) generating an electronic image of the 2-D separation of **polypeptides** of the two sources;

(c) warping one of the electronic images of the 2-D separation of **polypeptides** to the other image;

(d) analyzing the two 2-D separation of **polypeptides** of the sources to determine **polypeptide** spots common to both tissues;

(e) confirming commonality of at least a portion of the **polypeptide** spots common in both the two 2-D separation of **polypeptides** ;

(f) recording in a **database polypeptide** spots common to both tissues as being the same in response to positive confirmation of the portion of the spots common to both 2D separation of **polypeptides** ;

(g) analyzing **polypeptide** spots not common to both 2-D separations; and

(h) recording in the **database** results of the analyzing the **polypeptide** spots not common to both 2-D separations;

(7) identifying a **polypeptide** in a sample from an individual of a randomly breeding population comprising:

(a) characterizing the **polypeptide** by isoelectric point and **molecular** weight;

(b) identifying tissues of the subject where the **polypeptide** is found to yield distinguishing parameters of the **polypeptide** comprising isoelectric point, **molecular** weight and tissue distribution;

(c) **comparing** parameters with distinguishing parameters of previously tested **polypeptides** of a set; and

(d) determining whether a previously tested **polypeptide** has the parameters of the **polypeptide** ; and

(8) a data processing system for determining identity of an element (N+1) to N elements of a **database** contained in a storage medium comprising computer processing mechanism, data storage mechanism, and mechanism for processing data regarding **comparing** a parameter of the

(N+1) element with the parameter of the N elements of the **database** , where:

(a) the element is a **protein** or **polypeptide** ;

(b) processing data is **repeated** at least M times, where each M parameter is examined at each **iteration** (where M is at least 3) and when the (N+1) element does not have M identical parameters of N element(s), the data storage mechanism adds data of the (N+1) element and of the M parameters to the **database** to produce a new **database** comprising (N+1) elements;

(c) the **database** comprises **database** elements corresponding to **proteins** in tissues obtained from a selected organism; and

(d) a difference in abundance of the candidate **protein** marker identifies the candidate **protein** marker as a **protein** marker for the condition.

USE - For organizing **database** elements corresponding to **proteins** in tissue obtained from a selected organism, organelle, cell, tissue, organ, or population.

ADVANTAGE - The invention can measure the same **protein** in multiple different tissues. It can also measure the abundance of a **protein** at a particular **location** .

DESCRIPTION OF DRAWING(S) - The figure is a schematic block diagram showing the steps that form **part** of the analysis for **comparing proteins** of different tissues.

pp; 93 DwgNo 1/9

Title Terms: COMPUTER; READ; STRUCTURE; USEFUL; ORGANISE; **DATABASE** ; ELEMENT; CORRESPOND; **PROTEIN** ; TISSUE; OBTAIN; ORGANISM; COMPRISE; RECORD; PARAMETER; FIELD; **LOCATE** ; FIELD; ABUNDANT; FIELD

Derwent Class: B04; C07; D16; S03; T04

International Patent Class (Main): **G06F-017/60** ; **G06F-019/00** ; G06K-009/00

International Patent Class (Additional): B01D-057/02; C07K-014/00; G01N-033/48; G01N-033/50

File Segment: CPI; EPI

17/5/5 **(Item 5 from file: 350)**
DIALOG(R)File 350:Derwent WPIX

013844979
WPI Acc No: 2001-329192/200134
Related WPI Acc No: 2004-119084
XRAM Acc No: C01-101043
XRPX Acc No: N01-236924

Computer-based method of drug design that uses three-dimensional protein **structural models derived from** genetic **polymorphisms, useful for modifying existing drugs and identifying potential drug candidates**

Patent Assignee: QUEST DIAGNOSTICS INVESTMENTS INC (QUES-N); STRUCTURAL BIOINFORMATICS INC (STRU-N)
Inventor: HESS P P; MAGGIO E T; RAMNARAYAN K
Number of Countries: 095 Number of Patents: 003
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 200135316	A2	20010517	WO 2000US30863	A	20001110	200134	B
AU 200117600	A	20010606	AU 200117600	A	20001110	200152	
EP 1228370	A2	20020807	EP 2000980321	A	20001110	200259	
			WO 2000US30863	A	20001110		

Priority Applications (No Type Date): US 2000704362 A 20001101; US 99438566 A 19991110
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200135316	A2	E	368	G06F-019/00	

Designated States (National): AE AG AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

AU 200117600	A		G06F-019/00	Based on patent WO 200135316
EP 1228370	A2	E	G01N-033/50	Based on patent WO 200135316

Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR

Abstract (Basic): WO 200135316 A2

NOVELTY - A computer-based method (M1) of drug design that uses three-dimensional (3-D) **protein** structural models derived from **genetic** polymorphisms, is new.

DETAILED DESCRIPTION - A computer-based method (M1) of drug design that uses three-dimensional (3-D) **protein** structural models derived from **genetic** polymorphisms, is new.

M1 comprises:

(a) obtaining more than one **amino acid** sequence of target **proteins** that are the product of a gene exhibiting **genetic** polymorphisms, where the sequences represent different **genetic** polymorphisms;

(b) generating 3-D **protein** structural variant models from the sequences; and

(c) based upon the structures of the 3-D models, designing drug candidates, modifying existing drugs, identifying potential drug candidates or identifying modifications of existing drugs based on predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants.

INDEPENDENT CLAIMS are also included for the following:

(1) a computer-based method (M2) of selecting drug therapies for patients based on **genetic** polymorphisms, comprising:

(a) step (a) and (b) of M1;

(b) computationally docking drug **molecules** with the target **protein** models;

(c) energetically refining the docked complexes;

(d) determining the binding interactions between the drug or potential 15 new drug candidate **molecules** and the models; and

(e) selecting drug therapies based on the drug or drugs that have

the most favorable binding interactions with the structural variant models;

(2) a computer-based method for predicting clinical responses in patients based on **genetic** polymorphisms, comprising:

(a) steps (a) and (b) of M1;

(b) building a relational **database** of **protein** structural variants derived based on **genetic** polymorphisms and observed clinical data associated with particular polymorphisms exhibited in the patients, where the **database** comprises 3-D **molecular** coordinates for the structural variant models, a **molecular graphics** interface for 3-D **molecular** structure visualization, computer functionality for **protein** sequence and structural analysis, **database searching** tools, and observed clinical data associated with the **genetic** polymorphisms, subject medical history and subject history associated with the **genetic** polymorphisms, obtaining a target **protein** structural variant based on the same gene associated with a polymorphism in a patient;

(c) generating a 3-D **protein** model based on the subject's gene sequence;

(d) screening/ **comparing** the 3-D model derived from the subject to the structures contained in the **database** by identifying structures in the **database** that are similar to the model derived from the subject and predicting a clinical outcome for the patient based on the clinical data associated with the identified structures;

(3) a computer-based method for designing therapeutic agents that are active **against** biological targets that have become drug resistant due to **genetic** mutations, comprising obtaining a first 3-D **protein** structural variant model of a target **protein** **against** which a given drug has biological activity, generating a second 3-D **protein** structural variant model of the target in which **genetic** mutations have occurred and **against** which the same drug is no longer biologically active, **comparing** the structures of the first and second model to identify structural differences, and performing structure-based drug design calculations in order to identify new drugs or modifications to the existing drug to bring about biological activity **against** the second model;

(4) a computer-based method for identifying compensatory mutations in a target **protein** , comprising obtaining the **amino** **acid** sequence of a target **protein** containing multiple **amino** **acid** mutations that is expressed in a patient, where the structure of a form of the target **protein** that responds to a particular drug, including the active site, has been structurally characterized, generating a 3-D structural model of the mutated **protein** ; **comparing** the structure of the mutated **protein** with the form of the **protein** that responds to the drug to identify structural differences and/or similarities arising from the mutations, **comparing** the biological activities of the drug **against** both the mutated **protein** and the form of the **protein** that responds to the drug to determine the effects of the mutations on drug response, and identifying the mutations in the **protein** that affect biological activity based on the **comparisons** ;

(5) a method (M3) for creating a 3-D structural polymorphism relational **database** , comprising obtaining one or more **amino** **acid** sequences of a target **protein** that is the product of a gene exhibiting a **genetic** polymorphism, where sequences represent different **genetic** polymorphisms, generating 3-D **protein** structural variant models from the sequences, energetically refining the models, evaluating the quality of the models, optionally obtaining associated clinical properties or data, and inputting the model and any associated properties and/or data into a relational **database** ;

(6) a **database** (D1) **created** by M3;

(7) a computer system, comprising a **database** containing data representative of the three dimensional structure of polymorphic variants of a drug target;

(8) a **database** (D2) comprising:

(a) sequences of nucleotides encoding a **protein** or its portions, where the **protein** comprises polymorphic variants and the portions encode a domain of the **protein** that comprises a site which binds to a drug candidate; andb) the coordinates of 3-D structures of the encoded

protein or its portions; and

(9) a **database** (D3) comprising the 115 nucleotide sequences defined in the specification that encode HIV protease or a portion of HIV reverse transcriptase.

USE - The computer-based method is useful for designing drug candidates, modifying existing drugs, identifying potential drug candidates or identifying modifications of existing drugs based on predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants. The method is also useful for understanding and overcoming drug resistance using the 3-D **protein** model structures resulting from multiple **genetic** polymorphisms or mutations in infectious agents e.g. HIV.

pp; 368 DwgNo 0/11

Title Terms: COMPUTER; BASED; METHOD; DRUG; DESIGN; THREE; DIMENSION; **PROTEIN** ; STRUCTURE; MODEL; DERIVATIVE; **GENETIC** ; POLYMORPH; USEFUL; MODIFIED; EXIST; DRUG; IDENTIFY; POTENTIAL; DRUG; CANDIDATE

Derwent Class: B04; D16; T01

International Patent Class (Main): G01N-033/50; **G06F-019/00**

International Patent Class (Additional): G01N-033/68

File Segment: CPI; EPI

17/5/7 **(Item 7 from file: 350)**
DIALOG(R)File 350:Derwent WPIX

012672954
WPI Acc No: 1999-479061/199940
XRAM Acc No: C99-140959

Identifying therapeutic polynucleotide targets from cells such as neoplastic cells, hyperproliferative cells, apoptotic cells, pathogen-infected cells or plant cells

Patent Assignee: GENZYME CORP (GENZ)
Inventor: ROBERTS B L; SHANKARA S

Number of Countries: 085 Number of Patents: 005
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 9937816	A1	19990729	WO 99US1463	A	19990125	199940	B
AU 9923391	A	19990809	AU 9923391	A	19990125	200001	
EP 1053349	A1	20001122	EP 99903346	A	19990125	200061	
			WO 99US1463	A	19990125		
JP 2002500896	W	20020115	WO 99US1463	A	19990125	200207	
			JP 2000528722	A	19990125		
AU 756357	B	20030109	AU 9923391	A	19990125	200320	

Priority Applications (No Type Date): US 98103230 P 19981005; US 98100436 P 19980126; US 9877853 P 19980313

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 9937816	A1	E	52	C12Q-001/68	
Designated States (National): AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
AU 9923391	A			C12Q-001/68	Based on patent WO 9937816
EP 1053349	A1	E		C12Q-001/68	Based on patent WO 9937816
Designated States (Regional): AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE					
JP 2002500896	W		52	C12Q-001/68	Based on patent WO 9937816
AU 756357	B			C12Q-001/68	Previous Publ. patent AU 9923391 Based on patent WO 9937816

Abstract (Basic): WO 9937816 A1

NOVELTY - A method for identifying a polynucleotide (PN) **fragment** of a gene conferring a selected phenotype to a sample cell, is new comprises:

DETAILED DESCRIPTION - The method (M1) comprises:

(a) obtaining a set of PNs representing gene expression in 2 or more sample cells;

(b) obtaining a set of PNs representing gene expression in one or more control cells; and

(c) identifying a unique PN representing a gene that is common to the 2 or more sample cells and differentially expressed in the sample cells **compared** to the control cell.

INDEPENDENT CLAIMS are also included for the following:

(1) a method for identifying one or more PNs corresponding to one or more secreted biological factors comprising:

(a) obtaining a set of PNs representing gene expression in one or more sample cells that secrete the factor;

(b) obtaining a set of PNs representing gene expression in one or more control cells that do not secrete the factor;

(c) identifying one or more unique PNs which are common to the sample cells, the unique PNs being absent or expressed at lower levels in the control cells;

(2) a method for identifying a therapeutic target comprising:

(a) obtaining a set of PNs representing gene expression in 2 or more sample cells;

(b) obtaining a set of PNs representing gene expression in one or

more control cells; and

(c) identifying a unique PN representing a gene that is common to the 2 or more sample cells and differentially expressed in the sample cells **compared** to the control cell;

(3) a method of **creating** a **database** of PN data resulting from processing cell samples comprising:

(a) transferring sequence records that correspond to PNs obtained from a sample of cells electronically to a computer processor and creating a data raw file containing observed PN abundances related to the samples; and

(b) **creating** a **compare data file** by combining the data raw file with other data raw files, the other data raw files having been created from other samples; where the **compare data file** contains records from the data raw files, the data having been normalized to indicate percentage of sample for the number of occurrences of a PN in each of samples from the cells;

(4) a system for identifying selected PN records comprising:

(a) a digital computer;

(b) a **database** coupled to the computer;

(c) a **database** coupled to a **database** server having data stored in it, the data comprising records of data combined from PN raw files, the data having been normalized to indicate percentage of sample for a number of occurrences of a same tag in each sample of the samples; and

(d) a code mechanism for applying queries based upon a desired selection criteria to the **data file** in the **database** to produce reports of PN records which **match** the desired selection criteria;

(5) a method for identifying selected PN records from a **database** , using a computer having a processor, memory, display, input/output devices, comprising:

(a) providing a **database** coupled to the computer having data stored in it the data comprising representations of data combined from PN raw files, the data having been normalized to indicate percentage of sample for a number of occurrences of a same PN in each of the samples; and

(b) using a code mechanism for applying queries based upon a desired selection criteria to the **data file** in the **database** to produce reports of PN records which **match** the desired selection criteria.

USE - The methods can be used with sample cells such as neoplastic cells, drug-resistant neoplastic cells, neoplastic cells which promote angiogenesis, de-differentiated cells, differentiated cells, apoptotic cells, hyperproliferative cells, cells infected with a pathogen, drug-resistant cells infected with a pathogen or plant cells. The selected phenotype may be associated with e.g. **genetic** disease, altered metabolic activity, senescence, apoptosis, drug metabolism or allergic reaction. Antibodies **against proteins** encoded by the identified PNs, immune effectors or antigen presenting cells presenting the **protein** , can be used with a cytokine or a co-stimulatory **molecule** for the therapy of disorders, e.g. for inducing an immune response **against** a **polypeptide** associated with a neoplastic phenotype (all claimed).

pp; 52 DwgNo 0/0

Title Terms: IDENTIFY; THERAPEUTIC; POLYNUCLEOTIDE; TARGET; CELL; NEOPLASMS ; CELL; CELL; CELL; PATHOGEN; INFECT; CELL; PLANT; CELL

Derwent Class: B04; D16

International Patent Class (Main): C12Q-001/68

International Patent Class (Additional): C12N-015/09; G01N-033/50; G01N-033/53; **G06F-017/30**

File Segment: CPI

```
Set     Items   Description
S1     406994   MOLECULE? OR MOLECULAR OR PROTEIN? OR PEPTIDE? OR AMINO()A-
            CID? OR GENETIC? OR POLYPEPTIDE?
S2     185958   DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK? OR FILE?) -
            OR DB OR DBS OR DBMS OR RDB OR RDBM OR OODB?
S3    1814894   MATCH? OR COMPAR? OR QUERY OR QERIE? OR QUERYING OR SEARCH?
             OR LOCAT? OR FIND? OR SEEK?
S4     920653   REPEAT? OR ITERAT? OR REITERAT? OR AGAIN?
S5     545917   GRAPH? OR PARENT? OR INDEX OR INDICE? OR LIST? ?
S6    1382223   FRAGMENT? OR CLIQUE? OR PART OR PARTS OR PARTIAL OR SECTIO-
            N? OR STRING? OR SUBSTRING? OR MF OR MFS OR RESIDUE? OR CHAIN?
S7        120   S1(10N)S2(10N)S3(10N)S4(10N)S5(10N)S6
S8        153   S1(2N)S6(10N)S2(2N)(CREAT? OR MAKE? OR FILL? OR POPULAT? OR
               DEVELOP? OR BUILD?)
S9       1480   S2(4N)S1(4N)S6
S10         9   S7(S)S8
S11        30   S7(S)S9
S12        84   S8(S)S9
S13        12   (S10 OR S11 OR S12) AND IC=G06F?
S14        21   S10 OR S13
S15       744   S1(2N)S2(2N)S6
S16        83   S8 AND S15
S17        25   S16 NOT S12
S18         2   S17 AND IC=(G06F? OR H04L?)
S19        14   S18 OR S13
S20        24   S8(3N)(S4 OR S5)
S21         1   S20 AND IC=G06F?
S22        14   S21 OR S18 OR S13
S23        14   IDPAT (sorted in duplicate/non-duplicate order)
S24        14   IDPAT (primary/non-duplicate records only)
File 348:EUROPEAN PATENTS 1978-2004/Apr W02
         (c) 2004 European Patent Office
File 349:PCT FULLTEXT 1979-2002/UB=20040415,UT=20040408
         (c) 2004 WIPO/Univentio
```

24/5,K/2 (Item 2 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

01100428 **Image available**
SEARCHABLE MOLECULAR DATABASE
BASE DE DONNEES MODULAIRE CONSULTABLE
Patent Applicant/Assignee:
CRESSET BIOMOLECULAR DISCOVERY LIMITED, Spierlla Buildings, Suite 203, Bridge Road, Letchworth, Hertfordshire SG6 4ET, GB, GB (Residence), GB (Nationality), (For all designated states except: US)
Patent Applicant/Inventor:
ASHWORTH Philip Anthony, c/o Cresset Biomolecular Discovery Limited, Spirella Building, Suite 203, Bridge Road, Letchworth, Hertfordshire SG6 4ET, GB, GB (Residence), GB (Nationality), (Designated only for: US)
CHEESERIGHT Tim, c/o Cresset Biomolecular Discovery Limited, Spierlla Buildings, Suite 203, Bridge Road, Letchworth, Hertfordshire SG6 4ET, GB, GB (Residence), GB (Nationality), (Designated only for: US)
MACKEY Mark Denis, c/o Cresset Biomolecular Discovery Limited, Spierlla Buildings, Suite 203, Bridge Road, Letchworth, Hertfordshire SG6 4ET, GB, GB (Residence), GB (Nationality), (Designated only for: US)
VINTER Jeremy Gilbert, c/o Cresset Biomolecular Discovery Limited, Spierlla Buildings, Suite 203, Bridge Road, Letchworth, Hertfordshire SG6 4ET, GB, GB (Residence), GB (Nationality), (Designated only for: US)
Legal Representative:
HAINES Miles John (et al) (agent), D Young & Co, 21 New Fetter Lane, London EC4A 1DA, GB,
Patent and Priority Information (Country, Number, Date):
Patent: WO 200423337 A1 20040318 (WO 0423337)
Application: WO 2003GB3868 20030905 (PCT/WO GB03003868)
Priority Application: GB 200220790 20020906
Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NI NO NZ OM PG PH PL PT RO RU SC SD SE SG SK SL SY TJ TM TN TR TT TZ UA UG US UZ VC VN YU ZA ZM ZW
(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PT RO SE SI SK TR
(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW
(EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: **G06F-017/30**
International Patent Class: **G06F-017/50** ; **G06F-019/00**
Publication Language: English
Filing Language: English
Fulltext Availability:
Detailed Description
Claims
Fulltext Word Count: 12501

English Abstract
A computer system comprising a database (100) having a plurality of records is provided. Each record comprises a filed point representation representing field extrema for a conformation of a chemical structure. The database may include records for multiple conformations of the same chemical structure. Each record can have a searchable index of the filed point representation. In one embodiment the index is bit string. An indexing mechanism for generating an index, a searching mechanism for searching the database and a graphical user interface to enable a user to interface with the database (100) are also provided.

French Abstract
L'invention concerne un systeme informatique comprenant une base de donnees (100) qui possede une pluralite de fichiers. Chaque fichier comprend une representation de point de champ representant des extremites de champ pour une conformation d'une structure chimique. La base de

donnees comprend des fichiers pour des conformations multiples de la meme structure chimique. Chaque fichier peut presenter un index consultable de la representation de point de champ. Dans un mode de realisation, l'index est une chaine de bits. L'invention concerne egalement un mecanisme d'indexation permettant de generer un index, un mecanisme de recherche permettant de consulter la base de donnees et une interface graphique utilisateur permettant a un utilisateur d'interagir avec la base de donnees (100).

Legal Status (Type, Date, Text)
Publication 20040318 A1 With international search report.
Publication 20040318 A1 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

Main International Patent Class: **G06F-017/30**
International Patent Class: **G06F-017/50** ...

... **G06F-019/00**
Fulltext Availability:
Detailed Description

Detailed Description
... key is generated. As the search proceeds, the search key is compared to the bit **string** of each **molecule** in the **database** . If a TRUE bit in the search key is not also set as TRUE in

24/5,K/3 (Item 3 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

01025688
METHODS AND DEVICES FOR PROTEOMICS DATA COMPLEXITY REDUCTION
PROCEDES ET DISPOSITIFS DE REDUCTION DE LA COMPLEXITE DE DONNEES PROTEOMIQUES

Patent Applicant/Assignee:
IRM LLC, c/o Sophia House, 48 Church Street, Hamilton, HM LX, BM, -- (Residence), US (Nationality), (For all designated states except: US)
Patent Applicant/Inventor:
BROCK Ansgar, 10847 Caminito Alto, San Diego, CA 92131, US, US (Residence), DE (Nationality), (Designated only for: US)
HORN David M, 860 Turquoise Street, #327, San Diego, CA 92109, US, US (Residence), US (Nationality), (Designated only for: US)
PETERS Eric C, 3381 Calle del Sur, Carlsbad, CA 92009, US, US (Residence), US (Nationality), (Designated only for: US)
Legal Representative:
QUINE Jonathan Alan (et al) (agent), Quine Intellectual Property Law Group, P.C., P.O. Box 458, Alameda, CA 94501, US,
Patent and Priority Information (Country, Number, Date):
Patent: WO 200354772 A1 20030703 (WO 0354772)
Application: WO 2002US35607 20021105 (PCT/WO US0235607)
Priority Application: US 2001332988 20011105; US 2002368342 20020327; US 2002385769 20020603; US 2002385364 20020603; US 2002385835 20020603; US 2002386915 20020605; US 2002410382 20020912
Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VC VN YU ZA ZM ZW
(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR
(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW
(EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: **G06F-019/00**
Publication Language: English
Filing Language: English
Fulltext Availability:
Detailed Description
Claims
Fulltext Word Count: 26108

English Abstract
Provided are methods and systems for identification of proteins using high mass accuracy mass spectrometry. Not only do high mass accuracy measurements provide greater confidence in protein identification assignments, but they also enable proteins to be identified with either less sequence coverage or fewer additional tandem MS experiments. In addition, high mass measurement accuracy optionally allows protein identifications to be made on the basis of the mass of a single peptide, providing higher-throughputs in the analysis of mixtures due to the significant decrease in time spent on additional tandem MS experiments. In addition, a concomitant time saving in the cross correlation process of mass spectral data with in silico digested databases would also be achieved.

French Abstract
L'invention concerne des procedes et des systemes destines a identifier des proteines a l'aide de spectrometrie de masse elevee, precise. Les mesures precises de masse elevee permettent une meilleure confiance dans les attributions d'identification de proteines mais elles permettent aussi d'identifier des proteines, soit avec une moindre couverture de sequence, soit avec moins d'experiences supplementaires de spectrometrie de masse en tandem. En outre, la mesure precise de masse elevee permet, eventuellement, de realiser des identifications de proteines reposant sur la masse d'un seul peptide, autorisant une plus grande productivite dans l'analyse de melanges en raison du raccourcissement du temps passe sur

des experiences supplementaires de spectrometrie de masse en tandem. On realise aussi une economie de temps concomitante dans le processus de correlation entre des donnees spectrales de masse et des bases de donnees de digestion in silico.

Legal Status (Type, Date, Text)
Publication 20030703 A1 With international search report.

Main International Patent Class: **G06F-019/00**
Fulltext Availability:
Detailed Description

Detailed Description
... each mass in the list of theoretical masses corresponds to one and only one unique **peptide** sequence). In this embodiment, correlation of an experimental peak with a unique mass from the...

...The data complexity reduction methods of the present invention can optionally be performed in an **iterative** manner, to further assign the unidentified MS peaks based upon information gleaned from the previous round of analysis. In this embodiment, after identification of one or more **parent protein** sequences (for example, by correlating an MS peak with a unique theoretical mass), the first database of identified proteins is regenerated to include the newly identified **parent protein** sequences (e.g., additional member proteins). Additional in silico **peptide fragments** are generated from the information in the updated first **database** , and the corresponding (unique and/or non-unique) theoretical masses are **again compared** to the **list** of mass peaks for the sample, to further reduce the number of unidentified MS peaks and to possibly correlate unassigned MS peaks to further additional **parent** proteins. The steps of regenerating the **list** of **parent** proteins, calculating theoretical masses for component peptides, and correlating the list to the remaining unidentified MS peaks is optionally **repeated** until no additional member proteins are identified.

[00161 Optionally, the member proteins in the sample (or proteolytically-cleaved **fragments** thereof) can be isotopically labeled prior to generating the mass list, to further assist in...

24/5,K/4 (Item 4 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

01018761 **Image available**
METHOD FOR MATCHING MOLECULAR SPATIAL PATTERNS
PROCEDE D'ADAPTATION DE MODELES MOLECULAIRES SPATIAUX
Patent Applicant/Assignee:
THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ILLINOIS, 352 Henry Administration, 506 South Wright Street, Urbana, IL 61801, US, US (Residence), US (Nationality), (For all designated states except: US)
Patent Applicant/Inventor:
BINKOWSKI Andrew T, 411 South Gables Boulevard, Wheaton, IL 60187, US, US (Residence), US (Nationality), (Designated only for: US)
ADAMIAN Larissa, 2037 Bunker Circle, Naperville, IL 60563, US, US (Residence), RU (Nationality), (Designated only for: US)
LIANG Jie, 133 Ashland Avenue, River Forest, IL 60305, US, US (Residence), US (Nationality), (Designated only for: US)
Legal Representative:
NOONAN Kevin E (agent), McDonnell Boehnen Hulbert & Berggoff, 300 South Wacker Drive, Chicago, IL 60606, US,
Patent and Priority Information (Country, Number, Date):
Patent: WO 200348724 A2-A3 20030612 (WO 0348724)
Application: WO 2002US38030 20021127 (PCT/WO US0238030)
Priority Application: US 2001333969 20011129; US 2001334689 20011130
Parent Application/Grant:
Related by Continuation to: US 2001334689 20011130 (CON); US 2001333969 20011129 (CON)
Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW
(EP) AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR
(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW
(EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: **G06F-017/11**
International Patent Class: **G06F-017/50**
Publication Language: English
Filing Language: English
Fulltext Availability:
Detailed Description
Claims
Fulltext Word Count: 21640

English Abstract
Structural alignment methods are described that compare the sequences of two or more structural features of molecules. The methods provide for a rigorous statistical analysis that can detect structural similarities in molecules regardless of the similarity in their primary sequences. Thus, the methods can be used to predict and explain functional properties of molecules from their three-dimensional conformation. The methods use databases of different structural features against which a query sequence can be searched. By combining the search results from the various databases, the functional properties of molecules can be predicted and serve as a basis for the efficient design of ligands, substrate analogues, inhibitors or pharmaceutical species thereof.

French Abstract
L'invention concerne des procedes d'alignement de structures consistant a comparer les sequences de deux ou de plusieurs caracteristiques de structures de molecules. Les procedes fournissent une analyse statistique rigoureuse capable de detecter des similitudes de structure dans les molecules, independamment de leurs sequences primaires. Les procedes peuvent donc etre utilises pour prevoir et expliquer les proprietes fonctionnelles de molecules a partir de leur configuration tridimensionnelle. Les procedes utilisent des bases de donnees de differentes caracteristiques de structures vis-a-vis desquelles une

sequence d'interrogation peut etre cherchee. En combinant les resultats des recherches provenant des diverses bases de donnees, les proprietes fonctionnelles des molecules peuvent etre prevues et servir de base pour la conception efficace de ligands, d'analogues de substrats, d'inhibiteurs ou d'especes pharmaceutiques de ceux-ci.

Legal Status (Type, Date, Text)
Publication 20030612 A2 Without international search report and to be republished upon receipt of that report.
Examination 20031023 Request for preliminary examination prior to end of 19th month from priority date
Search Rpt 20031127 Late publication of international search report
Republication 20031127 A3 With international search report.

Main International Patent Class: **G06F-017/11**
International Patent Class: **G06F-017/50**
Fulltext Availability:
Detailed Description

Detailed Description
... pocket. This process is
1 8
repeated for every pocket and void in the pvSoar **database** to **create** a new **database** of pocket and void signature of **amino acid residue** distributions (pvSoarD). The signature composition distributions can be compared to each other in any number...

24/5,K/11 (Item 11 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

00736204 **Image available**
METHOD AND SYSTEM FOR ARTIFICIAL INTELLIGENCE DIRECTED LEAD DISCOVERY THROUGH MULTI-DOMAIN CLUSTERING
PROCEDE ET SYSTEME DESTINES A LA DECOUVERTE DE POINTE ORIENTEE INTELLIGENCE ARTIFICIELLE A L'AIDE D'UN GROUPAGE MULTI-DOMAINE

Patent Applicant/Assignee:
BIOREASON INC, Suite 303, 150 Washington Avenue, Santa Fe, NM 87501, US, US (Residence), US (Nationality), (For all designated states except: US)
Patent Applicant/Inventor:
NICOLAOU Christodoulos A, 12 Aganippis Street, Limassol 3112, CY, CY (Residence), CY (Nationality), (Designated only for: US)
KELLEY Brian P, 480 Washington Avenue, Sante Fe, NM 87501, US, US (Residence), US (Nationality), (Designated only for: US)
NUTT Ruth F, 4 Colibri Tierra, Sante Fe, NM 87501, US, US (Residence), US (Nationality), (Designated only for: US)
BASSETT Susan I, 343 Calle Lorna Norte, Sante Fe, NM 87501, US, US (Residence), US (Nationality), (Designated only for: US)
Legal Representative:
AARONSON Lawrence H, McDonnell Boehnen Hulbert & Berghoff, Suite 3200, 300 South Wacker Drive, Chicago, IL 60606, US

Patent and Priority Information (Country, Number, Date):
Patent: WO 200049539 A1 20000824 (WO 0049539)
Application: WO 2000US4211 20000218 (PCT/WO US0004211)
Priority Application: US 99120701 19990219; US 99281990 19990329
Designated States: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE
(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG
(AP) GH GM KE LS MW SD SL SZ TZ UG ZW
(EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: **G06F-017/50**
Publication Language: English
Filing Language: English
Fulltext Availability:
Detailed Description
Claims
Fulltext Word Count: 35010

English Abstract
A system for analyzing a vast amount of data representative of chemical structure and activity information and concisely providing conclusions about structure-to-activity relationships. A computer may adaptively learn new substructure descriptors based on its analysis of the input data. The computer may then apply each substructure descriptor as a filter to establish new groups of molecules that match the descriptor. From each new group of molecules, the computer may in turn generate one or more additional new groups of molecules. A result of the analysis in an exemplary arrangement is a tree structure that reflects pharmacophoric information and efficiently establishes through lineage what effect on activity various chemical substructures are likely to have. The tree structure can then be applied as a multi-domain classifier, to help a chemist classify test compounds into structural subclasses.

French Abstract
L'invention concerne un systeme permettant d'analyser une grande quantite de donnees representant des structures chimiques et des informations d'activite, et donnant des conclusions concises concernant les relations structure-activite. Un ordinateur peut apprendre de maniere adaptative de nouveaux descripteur de sous-structures d'apres son analyse des donnees entrees. L'ordinateur peut ensuite appliquer chaque descripteur de sous-structure en tant que filtre en vue d'etablir de nouveaux groupes de

molecules correspondant au descripteur. A terme, l'ordinateur peut, a partir de chaque nouveau groupe, generer de nouveaux groupes de molecules supplementaires. Un resultat de l'analyse peut etre, par exemple, une structure arborescente refletant des informations pharmacophoriques et etablissant par des lignes les effets que differents produits chimiques sont susceptibles d'avoir sur l'activite. Ces structures arborescentes peuvent etre utilisees en tant que classeur multi-domaine, aux fins d'aider un chimiste a classer des composes test dans des sous-classes structurelles.

Legal Status (Type, Date, Text)

Publication	20000824	A1 With international search report.
Publication	20000824	A1 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
Examination	20001109	Request for preliminary examination prior to end of 19th month from priority date

Main International Patent Class: **G06F-017/50**
Fulltext Availability:
Detailed Description

Detailed Description
... sti-in(.") By way of example and without limitation, a useful system for representing chemical **molecules** in ASCII form is also provided by Daylight Chemical Information Systems, Inc. Daylight establishes a...

...can be used to specify substructures using rules that are straightfo rward extensions of SMILES **strings** . Additional inforination about Daylight SMARTS keys is provided at the Daylight web site indicated above.

According to Daylight, both SMILES and SMARTS **strings** employ atoms and bonds as fundamental svmbols, which can be used to specify the nodes and edges of a **molecule** 's **graph** and assism labels to the components of the **graph** . SMARTS **strings** are interpreted as patterns that can be **matched against** SMILES **string** representations of **molecules** , in the form of **database** queries for instance. Other examples of substructure representations include "MACCS" keys (i.e.. fi-aurnent-based keys for use in describing **molecules** , where MACCS stands for "the **Molecular** ACCess System) and other keys as defined by MDL Information Systems, Inc., for instance. (For...

24/5,K/12 (Item 12 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

00423320 **Image available**
SYSTEM AND METHOD FOR STRUCTURE-BASED DRUG DESIGN THAT INCLUDES ACCURATE PREDICTION OF BINDING FREE ENERGY
SYSTEME ET PROCEDE DE CONCEPTION RATIONNELLE DES MEDICAMENTS SUR LA BASE D'UNE STRUCTURE FAISANT INTERVENIR LA PREDICTION PRECISE DE L'ENERGIE LIBRE DE LIAISON
Patent Applicant/Assignee:
PRESIDENT AND FELLOWS OF HARVARD COLLEGE,
Inventor(s):
SHAKHNOVICH Eugene I,
DeWITTE Robert S,
Patent and Priority Information (Country, Number, Date):
Patent: WO 9813781 A1 19980402
Application: WO 97US17201 19970925 (PCT/WO US9717201)
Priority Application: US 96741866 19960926
Designated States: JP AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE
Main International Patent Class: **G06F-019/00**
Publication Language: English
Fulltext Availability:
Detailed Description
Claims
Fulltext Word Count: 11078

English Abstract
A system and method for providing improved de novo structure based drug design that include a method for more accurately predicting binding free energy. The system and method use a coarse graining model with corresponding knowledge based potential data to grow candidate molecules or ligands (108). In light of the present invention using the coarse graining model, the novel growth method (108) of the present invention uses a metropolis Monte Carlo selection process (218) which result in a low energy structure that is not necessarily the lowest energy structure, yet a better candidate (110) can result.

French Abstract
L'invention porte sur un systeme et un procede de conception rationnelle des medicaments sur la base d'une structure de novo, amelioree, et faisant intervenir un procede de prediction precise de l'energie libre de liaison. Ce systeme et ce procede utilisent un modele de granulation grossiere avec des donnees potentielles basees sur une connaissance correspondante de facon a developper des molecules ou ligand candidats (108). A la lumiere de la presente invention utilisant le modele de granulation grossiere, le nouveau procede de developpement moleculaire (108) met en oeuvre une methode de selection Metropolis Monte Carlo (218) qui donne lieu a une faible structure energetique, qui n'est pas necessairement la plus faible, mais qui, toutefois, permet d'obtenir un meilleur candidat (110).

Main International Patent Class: **G06F-019/00**
Fulltext Availability:
Detailed Description

Detailed Description
... greater detail subsequently.

It is known to use one of two methods to automatically search **databases** that contain large amounts of data relating to **fragments** that can be used for **building molecules** or ligands for developing lead candidates. A first method is the Geometric method that matches... functional groups. HOOK uses random placement of many copies of several functional fragments followed by **molecular** dynamics.

Multiple Start Monte Carlo methods also have been used as **fragmentjoining** methods. These methods conduct searches of **databases**

for **fragments** of a ligand to dock at the receptor site.
BUILDER software, uses a family of docked structures to provide an irregular lattice of controllable density...nearly I billion candidates of 5 functional groups -- 505 combinations. As the size of the **database** of **molecular fragments** increases, it is readily seen that the number of possible combinations will increase dramatically. As...

24/5,K/13 (Item 13 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

00386816 **Image available**

METHOD OF CREATING AND SEARCHING A MOLECULAR VIRTUAL LIBRARY USING VALIDATED MOLECULAR STRUCTURE DESCRIPTORS

PROCEDE POUR CREER UNE BIBLIOTHEQUE MOLECULAIRE VIRTUELLE ET PROCEDE POUR Y FAIRE DES RECHERCHES, EN UTILISANT DES DESCRIPTEURS VALIDES DE STRUCTURE MOLECULAIRE

Patent Applicant/Assignee:
PATTERSON David E,
CRAMER Richard D,
CLARK Robert D,
FERGUSON Allan M,
Inventor(s):
PATTERSON David E,
CRAMER Richard D,
CLARK Robert D,
FERGUSON Allan M,

Patent and Priority Information (Country, Number, Date):
Patent: WO 9727559 A1 19970731
Application: WO 97US1491 19970127 (PCT/WO US9701491)
Priority Application: US 96592132 19960126; US 96657147 19960603
Designated States: AU CA CN CZ HU IL JP KR NO PL US AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE
Main International Patent Class: **G06F-019/00**
Publication Language: English
Fulltext Availability:
Detailed Description
Claims
Fulltext Word Count: 125926

English Abstract
The problem of how to select out of a large chemically accessible universe molecules representative of the diversity of that universe is resolved by the discovery of a method to validate molecular structural descriptors. Using the validated descriptors, optimally diverse subsets (5) can be selected. In addition, from the universe, molecules with characteristics similar to a selected molecule can be identified (3). The validated descriptors also enable the generation of a huge virtual library of potential product molecules which could be formed by combinatorial arrangement of structural variations and cores. In this virtual library it is possible to search billions of possible product compounds in relatively short time frames.

French Abstract
Le probleme de la selection de molecules dans l'univers etendu des molecules chimiques possibles, dans toute sa diversite, est resolu par la decouverte d'un procede permettant de valider des descripteurs de structure moleculaire. En utilisant les descripteurs valides, on peut selectionner des sous-ensembles (5) diversifies de maniere optimale. En plus, on peut identifier (3) dans cet univers des molecules possedant des caracteristiques similaires a celles d'une molecule selectionnee. Les descripteurs valides permettent, egalement, de produire une bibliotheque virtuelle immense de molecules potentielles de produits qui peuvent etre formees par arrangement combinatoire de differentes structures et noyaux. Dans cette bibliotheque virtuelle, il est possible d'effectuer une recherche parmi des milliards de composes possibles de produits, en un temps relativement court.

Main International Patent Class: **G06F-019/00**
Fulltext Availability:
Detailed Description

Detailed Description
... of just the side chains (as was done with the topomeric CoMFA metric) of the **molecules** for the same 20 data sets. In Table 3 are shown the

Tanimoto fingerprint density ratios for the whole **molecule** and side **chain** Tanimoto metrics and the corresponding X' values for the 20 **data** sets.

TABLE 3
Patterson Plot Ratios and Associated X2
Col, I Col. 2 Col, 3...metric is more sensitive to the volume and shape of the space occupied by a **molecule** than is, for instance, either the side **chain** or whole **molecule** Tanimoto descriptor. Figure 12 provides an illustrative example of this feature drawn from the thiol...

```
Set     Items   Description
S1    6548042   MOLECULE? OR MOLECULAR OR PROTEIN? OR PEPTIDE? OR AMINO()A-
            CID? OR GENETIC? OR POLYPEPTIDE?
S2     905869   DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK? OR FILE?) -
            OR DB OR DBS OR DBMS OR RDB OR RDBM OR OODB?
S3   10011342   MATCH? OR COMPAR? OR QUERY OR QERIE? OR QUERYING OR SEARCH?
             OR LOCAT? OR FIND? OR SEEK?
S4    1781117   REPEAT? OR ITERAT? OR REITERAT? OR AGAIN?
S5    2566570   GRAPH? OR PARENT? OR INDEX OR INDICE? OR LIST? ?
S6    6084331   FRAGMENT? OR CLIQUE? OR PART OR PARTS OR PARTIAL OR SECTIO-
            N? OR STRING? OR SUBSTRING? OR MF OR MFS OR RESIDUE? OR CHAIN?
S7        614   S1(2N)S6(2N)S2
S8        128   S1 AND S2 AND S3 AND S4 AND S5 AND S6
S9      34658   S2(2N)(CREAT? OR POPULAT? OR FILL? OR DEVELOP? OR BUILD?)
S10         6   S8 AND S9
S11        19   S7 AND S9
S12         4   S7 AND S8
S13        91   S1(3N)S6 AND S9
S14        49   S13 AND (S3 OR S5)
S15        25   S13 AND S4
S16        82   S10 OR S11 OR S12 OR S14 OR S15
S17        68   RD (unique items)
S18        43   S17 NOT PY>2000
S19        43   S18 NOT PD=20001117:20021117
S20        43   S19 NOT PD=20021117:20040501
File   2:INSPEC 1969-2004/Apr W2
         (c) 2004 Institution of Electrical Engineers
File   6:NTIS 1964-2004/Apr W3
         (c) 2004 NTIS, Intl Cpyrght All Rights Res
File   8:Ei Compendex(R) 1970-2004/Apr W2
         (c) 2004 Elsevier Eng.  Info. Inc.
File  34:SciSearch(R) Cited Ref Sci 1990-2004/Apr W2
         (c) 2004 Inst for Sci Info
File  35:Dissertation Abs Online 1861-2004/Mar
         (c) 2004 ProQuest Info&Learning
File  65:Inside Conferences 1993-2004/Apr W3
         (c) 2004 BLDSC all rts. reserv.
File  94:JICST-EPlus 1985-2004/Apr W1
         (c)2004 Japan Science and Tech Corp(JST)
File  95:TEME-Technology & Management 1989-2004/Apr W1
         (c) 2004 FIZ TECHNIK
File  99:Wilson Appl. Sci & Tech Abs 1983-2004/Mar
         (c) 2004 The HW Wilson Co.
File 144:Pascal 1973-2004/Apr W2
         (c) 2004 INIST/CNRS
File 202:Info. Sci. & Tech. Abs. 1966-2004/Feb 27
         (c) 2004 EBSCO Publishing
File 233:Internet & Personal Comp. Abs. 1981-2003/Sep
         (c) 2003 EBSCO Pub.
```

20/5/1 (Item 1 from file: 2)
DIALOG(R)File 2:INSPEC

5647405 INSPEC Abstract Number: A9717-3310-015

Title: Adiabatic semi-empirical parametric method for computing electronic-vibrational spectra of complex molecules. 1. Polyenes and diphenylpolyenes

Author(s): Baranov, V.I.; Gribov, L.A.; Djenjer, V.O.; Zelent'sov, D.Yu.
Author Affiliation: Vernadsky Inst. of Geochem. & Anal. Chem., Acad. of Sci., Moscow, Russia

Journal: Journal of Molecular Structure vol.407, no.2-3 p.177-98
Publisher: Elsevier,
Publication Date: 30 May 1997 Country of Publication: Netherlands
CODEN: JMOSB4 ISSN: 0022-2860
SICI: 0022-2860(19970530)407:2/3L.177:ASEP;1-I
Material Identity Number: J126-97014
U.S. Copyright Clearance Center Code: 0022-2860/97/$17.00
Document Number: S0022-2860(96)09611-1
Language: English Document Type: Journal Paper (JP)
Treatment: Theoretical (T)

Abstract: A parametric semi-empirical method for the calculation of the vibrational structure of the electronic spectrum and the determination of the parameters of the molecular excited state potential surface has been developed. The method is based on the adiabatic molecular model and is unique for all sets of parameters of the excited states (first and second derivatives of the matrix of coulombic and resonant one-electron integrals with respect to the internal coordinates). Simplified analytical expressions for the changes in the molecular potential surfaces on excitation, which account only for the first-order terms, are obtained. It is shown that the parameters possess distinct local properties and may be transferred in a homologous series of molecules. The number of most significant parameters, sufficient to describe the molecular model adequately and to obtain satisfactory quantitative results, is very small. Calculations of geometry changes and vibronic spectra for some polyene and diphenylpolyene molecules using only two parameters show good quantitative agreement with experimental data. It is possible to **create** a special **data bank** of **molecular fragments** for vibronic spectroscopy with relatively small structural groups (e.g. H>C= for polyenes and related compounds) and to use it to compute the excited state properties of complex molecules and their vibronic spectra employing the suggested parametric method. (38 Refs)

Subfile: A
Descriptors: excited states; polymers; potential energy surfaces; spectra ; vibrational states
Identifiers: semiempirical parametric method; electronic-vibrational spectra; complex molecules; polyenes; diphenylpolyenes; parametric semiempirical method; vibrational structure; electronic spectrum; molecular excited state potential surface; adiabatic molecular model; coulombic one-electron integrals; resonant one-electron integrals; internal coordinates; analytical expressions; molecular potential surfaces; molecular model; vibronic spectra; **molecular fragments data bank** ; vibronic spectroscopy; excited state properties
Class Codes: A3310G (Vibrational analysis (molecular spectra)); A3620K (Electronic structure and spectra of macromolecules); A3150 (Excited states of atoms and molecules)

20/5/2 **(Item 2 from file: 2)**
DIALOG(R)File 2:INSPEC

5300365 INSPEC Abstract Number: A9615-3120E-001, C9608-7320-014

Title: A high-resolution shape-fragment MEDLA database for toxicological shape analysis of PAHs

Author(s): Mezey, P.G.; Zimpel, Z.; Warburton, P.; Walker, P.D.; Irvine, D.G.; Dixon, D.G.; Greenberg, B.

Author Affiliation: Dept. of Chem., Saskatchewan Univ., Saskatoon, Sask., Canada

Journal: Journal of Chemical Information and Computer Sciences vol.36, no.3 p.602-11

Publisher: ACS,

Publication Date: May-June 1996 Country of Publication: USA

CODEN: JCISD8 ISSN: 0095-2338

SICI: 0095-2338(199605/06)36:3L.602:HRSF;1-5

Material Identity Number: J263-96003

U.S. Copyright Clearance Center Code: 0095-2338/96/1636-0602$12.00/0

Language: English Document Type: Journal Paper (JP)

Treatment: Practical (P)

Abstract: A new, high-resolution shape-fragment **database** has been **developed** for computing ab initio quality molecular electron densities for polyaromatic hydrocarbons (PAHs) which play a significant role as toxicants in the environment. Using the new PAH electron density **fragment database** and the **Molecular** Electron Density Lego Assembler (MEDLA) method, one can generate detailed and reliable electron densities for virtually any of the PAH molecules. Accurate electron density shape representations for these molecules is essential in the study of detailed shape-toxicity correlations. One of our goals is to investigate the potential of detailed molecular shape analysis as a predictive tool in toxicological risk assessment. In this study we report the results of the first phase of the study: the construction and testing of a high quality shape-fragment database for PAHs. (45 Refs)

Subfile: A C

Descriptors: ab initio calculations; chemistry computing; database management systems; molecular electronic states; organic compounds

Identifiers: high-resolution shape-fragment MEDLA database; toxicological shape analysis; ab initio quality molecular electron densities; polyaromatic hydrocarbons; toxicants; Molecular Electron Density Lego Assembler; electron density shape representations; shape-toxicity correlations; toxicological risk assessment

Class Codes: A3120E (Ab initio LCAO and GO SCF calculations (atoms and molecules)); C7320 (Physics and chemistry computing); C6160 (Database management systems (DBMS))

20/5/7 **(Item 4 from file: 34)**
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci

08513220 Genuine Article#: 295AX Number of References: 41
Title: TOP: a new method for protein structure comparisons **and similarity** searches
Author(s): Lu GG (REPRINT)
Corporate Source: LUND UNIV,DEPT MOL BIOPHYS, BOX 124/S-22100 LUND//SWEDEN/ (REPRINT); KAROLINSKA INST,DEPT MED BIOCHEM & BIOPHYS, DIV MOL STRUCT BIOL/S-17177 STOCKHOLM//SWEDEN/
Journal: JOURNAL OF APPLIED CRYSTALLOGRAPHY, 2000, V33, 1 (FEB), P176-183
ISSN: 0021-8898 Publication date: 20000200
Publisher: MUNKSGAARD INT PUBL LTD, 35 NORRE SOGADE, PO BOX 2148, DK-1016 COPENHAGEN, DENMARK
Language: English Document Type: ARTICLE
Geographic Location: SWEDEN
Subfile: CC PHYS--Current Contents, Physical, Chemical & Earth Sciences
Journal Subject Category: CRYSTALLOGRAPHY
Abstract: In order to facilitate the three-dimensional structure **comparison** of proteins, software for making **comparisons** and **searching** for similarities to protein structures in **databases** has been **developed** . The program identifies the residues that share similar positions of both main-chain and side- **chain** atoms between two **proteins** . The unique functions of the software also include database processing via Internet- and Web-based servers for different types of users. The developed method and ifs friendly user interface copes with many of the problems that frequently occur in protein structure **comparisons** , such as detecting structurally equivalent residues, misalignment caused by coincident **match** of C-alpha atoms, circular sequence permutations, tedious repetition of access, maintenance of the most recent database, and inconvenience of user interface. The program is also designed to cooperate with other tools in structural bioinformatics, such as the 3DB Browser software [Prilusky (1998), Protein Data Bank Q. Newslett. 54, 3-4] and the SCOP database [Murzin, Brenner, Hubbard & Chothia (1995). J. Mel. Biol. 247, 536-540], for convenient molecular modelling and protein structure analysis. A similarity ranking score of 'structure diversity' is proposed in order to estimate the evolutionary distance between proteins based on the **comparisons** of their three-dimensional structures. The function of the program has been utilized as a part of an automated program for multiple protein structure alignment. In this paper, the algorithm of the program and results of systematic tests are presented and discussed.
Identifiers--KeyWord Plus(R): CRYSTAL-STRUCTURE; FLAVOPROTEIN REDUCTASES; CIRCULAR PERMUTATION; SECONDARY STRUCTURE; DATA-BANK; FAMILY; MOTIFS; RESOLUTION; ALIGNMENT; DOMAINS
Cited References:
*COLL COMP PROJ 4, 1994, V50, P760, ACTA CRYSTALLOGR D
ALEXANDROV NN, 1996, V25, P354, PROTEINS
BERNSTEIN FC, 1977, V112, P535, J MOL BIOL
BRENNER SE, 1998, V95, P6073, P NATL ACAD SCI USA
CHOI HK, 1997, V27, P345, PROTEINS
CORREL CC, 1992, V258, P1064, SCIENCE
CORRELL CC, 1993, V2, P2112, PROTEIN SCI
DOBRITZSCH D, 1998, V273, P20196, J BIOL CHEM
ENROTH, 1998, THESIS KAROLINSKA I
ENROTH C, 1998, V6, P605, STRUCTURE
FISCHER D, 1992, V9, P769, J BIOMOL STRUCT DYN
GERSTEIN M, 1998, V7, P455, PROTEIN SCI
GIBRAT JF, 1996, V6, P377, CURR OPIN STRUC BIOL
GRINDLEY HM, 1993, V229, P707, J MOL BIOL
HOLM L, 1993, V233, P123, J MOL BIOL
HOLM L, 1996, V34, P206, NUCLEIC ACIDS RES
HOLM L, 1994, V19, P165, PROTEINS
HUANG W, 1998, V7, P1183, EMBO J
HUBER R, 1965, V19, P353, ACTA CRYSTALLOGR
JIA J, 1996, V4, P715, STRUCTURE

JONES TA, 1991, V47, P110, ACTA CRYSTALLOGR A
KABSCH W, 1983, V22, P2577, BIOPOLYMERS
KLEYWEGT GJ, 1997, V277, P525, METHOD ENZYMOL
LASKOWSKI RA, 1993, V26, P283, J APPL CRYSTALLOGR
LIEPINSH E, 1997, V4, P975, NAT STRUCT BIOL
LINDQVIST Y, 1997, V7, P422, CURR OPIN STRUC BIOL
LU G, 1996, V78, P10, PROTEIN DATA BANK Q
LU GG, 1994, V2, P809, STRUCTURE
MATTHEWS BW, 1985, V115, P397, METHOD ENZYMOL
MEDEJ T, 1995, V23, P356, J PROTEIN STRUCT FUN
MITCHELL EM, 1990, V212, P151, J MOL BIOL
MIZUGUCHI K, 1998, V7, P2469, PROTEIN SCI
MURZIN AG, 1998, V5, P101, NAT STRUCT BIOL
ORENGO CA, 1992, V14, P139, PROTEINS
ORENGO CA, 1997, V5, P1093, STRUCTURE
PRILUSKY J, 1998, V84, P3, PROTEIN DATA BANK Q
ROSSMANN MG, 1975, V105, P75, J MOL BIOL
SALI A, 1990, V212, P403, J MOL BIOL
SOWDHAMINI R, 1996, V1, P209, FOLD DES
SUBBIAH S, 1993, V3, P141, CURR BIOL
VRIEND G, 1991, V11, P52, PROTEINS

20/5/14 **(Item 11 from file: 34)**
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci

05695197 Genuine Article#: WQ624 Number of References: 42
Title: Similarity searching **in files of three-dimensional chemical structures: Representation and** searching **of** molecular **electrostatic potentials using field-** graphs
Author(s): Thorner DA; Willett P (REPRINT) ; Wright PM; Taylor R
Corporate Source: UNIV SHEFFIELD,KREBS INST BIOMOLEC RES/SHEFFIELD S10 2TN/S YORKSHIRE/ENGLAND/ (REPRINT); UNIV SHEFFIELD,KREBS INST BIOMOLEC RES/SHEFFIELD S10 2TN/S YORKSHIRE/ENGLAND/; UNIV SHEFFIELD,DEPT INFORMAT STUDIES/SHEFFIELD S10 2TN/S YORKSHIRE/ENGLAND/; ZENECA AGROCHEM,JEALOTTS HILL RES STN/BRACKNELL RG12 6EY/BERKS/ENGLAND/
Journal: JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN, 1997, V11, N2 (MAR), P 163-174
ISSN: 0920-654X Publication date: 19970300
Publisher: ESCOM SCI PUBL BV, PO BOX 214, 2300 AE LEIDEN, NETHERLANDS
Language: English Document Type: ARTICLE
Geographic Location: ENGLAND
Subfile: CC LIFE--Current Contents, Life Sciences
Journal Subject Category: BIOCHEMISTRY & MOLECULAR BIOLOGY
Abstract: This paper reports a method for the identification of those **molecules** in a **database** of rigid 3D structures with **molecular** electrostatic potential (MEP) grids that are most similar to that of a user-defined target **molecule** . The most important features of an MEP grid are encoded infield- **graphs** , and a target **molecule** is **matched against** a **database molecule** by a **comparison** of the corresponding field- **graphs** . The **matching** is effected using a maximal common subgraph isomorphism algorithm, which provides an alignment of the target **molecule** 's field- **graph** with those of each of the **database molecules** in turn. These alignments are used in the second stage of the **search** algorithm to calculate the intermolecular MEP similarities. Several different ways of generating field- **graphs** are evaluated, in terms of the effectiveness of the resulting similarity measures and of the associated computational costs. The most appropriate procedure has been implemented in an operational system that **searches** a corporate **database** , containing ca. 173 000 3D structures.
Descriptors--Author Keywords: **clique** -detection algorithm ; **database searching** ; field- **graph** ; **molecular** electrostatic potential ; similarity **searching**
Identifiers--KeyWord Plus(R): 3D **DATABASE** ; DRUG DESIGN; SUBSTRUCTURES; PROGRAM
Research Fronts: 95-1590 001 (**MOLECULAR** SIMILARITY; AB-INITIO QUALITY ELECTRON-DENSITIES FOR **PROTEINS** ; SHAPE GROUP-ANALYSIS)
95-4654 001 (3-DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY-RELATIONSHIPS; RECEPTOR SURFACE MODELS; **COMPARATIVE MOLECULAR** -FIELD ANALYSIS (COMFA); DRUG DISCOVERY)
Cited References:
ASH JE, 1991, CHEM STRUCTURE SYSTE
BRINT AT, 1987, V27, P152, J CHEM INF COMP SCI
BRINT AT, 1988, V2, P311, J COMPUT AID MOL DES
BRON C, 1973, V16, P575, COMMUN ACM
BURES MG, 1991, V5, P323, J COMPUT AID MOL DES
BURES MG, 1990, V3, P673, TETRAHEDRON COMP MET
BURES MG, 1994, V21, P467, TOP STEREOCHEM
BURT C, 1990, V11, P1139, J COMPUT CHEM
CARBO R, 1980, V17, P1185, INT J QUANTUM CHEM
CLARK DE, 1992, V10, P194, J MOL GRAPHICS
CRAMER RD, 1988, V110, P5959, J AM CHEM SOC
CRAMER RD, 1973, V17, P533, J MED CHEM
CRANDELL CW, 1983, V23, P186, J CHEM INF COMP SCI
GOOD AC, 1992, V32, P188, J CHEM INF COMP SCI
GOOD AC, 1992, V6, P513, J COMPUT AID MOL DES
GOOD AC, 1992, V10, P144, J MOL GRAPHICS
HAGADONE TR, 1992, V32, P515, J CHEM INF COMP SCI
HARAKI KS, 1990, V3, P565, TETRAHEDRON COMPUT M

HERMANN RB, 1991, V5, P511, J COMPUT AID MOL DES
HO CMW, 1993, V7, P3, J COMPUT AID MOL DES
HODGKIN EE, 1987, V14, P105, INT J QUANTUM CHEM
JOHNSON MA, 1990, CONCEPTS APPL MOL SI
KEARSLEY SK, 1990, V3, P615, TETRAHEDRON COMPUT M
LEVI G, 1972, V9, P341, CALCOLO
MANAUT F, 1991, V5, P371, J COMPUT AID MOL DES
MARSHALL GR, 1979, V112, P205, ACS SYM SER
MARTIN YC, 1993, V7, P83, J COMPUT AID MOL DES
MARTIN YC, 1992, V35, P2145, J MED CHEM
MEURICE N, IN PRESS HELV CHIM A
MILNE GWA, 1994, V34, P1219, J CHEM INF COMP SCI
PEPPERRELL CA, 1991, V33, P97, PESTIC SCI
PETKE JD, 1993, V14, P928, J COMPUT CHEM
REYNOLDS CA, 1992, V11, P34, QUANT STRUCT-ACT REL
RICHARD AM, 1991, V12, P959, J COMPUT CHEM
SANZ F, 1993, V7, P337, J COMPUT AID MOL DES
SNEATH PHA, 1973, NUMERICAL TAXONOMY P
STEWART JJP, 1990, V4, P1, J COMPUT AID MOL DES
THORNER DA, 1996, V36, P900, J CHEM INF COMP SCI
TURNER DB, 1995, V3, P101, SAR QSAR ENV RES
VANGEERESTEIN VJ, 1990, V3, P595, TETRAHEDRON COMPUT M
WILLETT P, 1995, V8, P290, J MOL RECOGNIT
WILLETT P, 1991, 3 DIMENSIONAL CHEM S

20/5/25 (Item 22 from file: 34)
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci

02956942 Genuine Article#: MT177 Number of References: 42
Title: SOPM - A SELF-OPTIMIZED METHOD FOR PROTEIN SECONDARY STRUCTURE PREDICTION
Author(s): GEOURJON C; DELEAGE G
Corporate Source: INST BIOL & CHEM PROT,CNRS,UPR 412,7 PASSAGE VERCORS/F-69367 LYON 07//FRANCE/; INST BIOL & CHEM PROT,CNRS,UPR 412/F-69367 LYON07//FRANCE/
Journal: PROTEIN ENGINEERING, 1994, V7, N2 (FEB), P157-164
ISSN: 0269-2139
Language: ENGLISH Document Type: ARTICLE
Geographic Location: FRANCE
Subfile: SciSearch; CC LIFE--Current Contents, Life Sciences
Journal Subject Category: BIOCHEMISTRY & MOLECULAR BIOLOGY
Abstract: A new method called the self-optimized prediction method (SOPM) has been developed to improve the success rate in the prediction of the secondary structure of proteins. This new method has been checked **against** an updated release of the Kabsch and Sander **database** , ' **DATABASE** .DSSP', comprising 239 **protein chains** . The first step of the SOPM is to **build** sub- **databases** of protein sequences and their known secondary structures drawn from 'DATABASE.DSSP', by (i) making binary **comparisons** of all protein sequences and (ii) taking into account the prediction of structural classes of proteins. The second step is to submit each protein of the sub-database to a secondary structure prediction using a predictive algorithm based on sequence similarity. The third step is to **iteratively** determine the predictive parameters that optimize the prediction quality on the whole sub-database. The last step is to apply the final parameters to the **query** sequence. This new method correctly predicts 69% of amino acids for a three-state description of the secondary structure (alpha helix, beta sheet and coil) in the whole database (46 011 amino acids). The correlation coefficients are C-alpha = 0.54, C-beta = 0.50 and C-c = 0.48. Root mean square deviations of 10% in the secondary structure content are obtained. Implications for the users are drawn so as to derive an accuracy at the amino acid level and provide the user with a guide for secondary structure prediction. The SOPM method is available by anonymous ftp to ibcp.fr.
Descriptors--Author Keywords: AMINO ACID SEQUENCE ; HOMOLOGY MODELING ; PROTEIN STRUCTURE ; SECONDARY STRUCTURE PREDICTION
Identifiers--KeyWords Plus: AMINO-ACID-SEQUENCE; NEURAL NETWORK; GLOBULAR PROTEINS; JOINT PREDICTION; ALGORITHM; CONFORMATION; ALIGNMENT; IMPROVEMENTS; INFORMATION; HOMOLOGIES
Research Fronts: 92-3995 003 (PROTEIN SECONDARY STRUCTURE; FUNCTIONAL TOPOGENIC DOMAINS; ALPHA-HELIX PREDICTION)
92-0078 002 (PROTEIN SECONDARY STRUCTURE; ARTIFICIAL NEURAL NETWORKS; ALPHA-HELIX PREDICTION)
Cited References:
BIOU V, 1988, V2, P185, PROTEIN ENG
BOSCOTT PE, 1993, V6, P261, PROTEIN ENG
BURGESS AW, 1974, V12, P239, ISRAEL J CHEM
CHOU PY, 1978, V47, P45, ADV ENZYMOL
CHOU PY, 1974, V13, P222, BIOCHEMISTRY-US
DELEAGE G, 1989, P587, PREDICTION PROTEIN S
DELEAGE G, 1987, V1, P289, PROTEIN ENG
FASMAN GD, 1989, P194, PREDICTION PROTEIN S
FENG DF, 1990, V183, P375, METHOD ENZYMOL
FINKELSTEIN AV, 1971, V62, P613, J MOL BIOL
GARNIER J, 1990, V72, P513, BIOCHIMIE
GARNIER J, 1991, V7, P133, COMPUT APPL BIOSCI
GARNIER J, 1978, V120, P97, J MOL BIOL
GEOURJON C, 1993, V9, P87, COMPUT APPL BIOSCI
GEOURJON C, 1991, V9, P188, J MOL GRAPHICS
GIBRAT JF, 1987, V198, P425, J MOL BIOL
HOLLEY LH, 1989, V86, P152, P NATL ACAD SCI USA
KABAT EA, 1973, V70, P1473, P NATL ACAD SCI USA

KABSCH W, 1983, V22, P2577, BIOPOLYMERS
KABSCH W, 1983, V155, P179, FEBS LETT
KABSCH W, 1984, V81, P1075, P NATL ACAD SCI USA
KNELLER DG, 1990, V214, P171, J MOL BIOL
LEVIN JM, 1988, V955, P283, BIOCHIM BIOPHYS ACTA
LEVIN JM, 1986, V205, P303, FEBS LETT
LIM VI, 1974, V88, P857, J MOL BIOL
MATTHEWS BW, 1975, V405, P442, BIOCHIM BIOPHYS ACTA
MUSKAL SM, 1992, V225, P713, J MOL BIOL
NAGANO K, 1977, V109, P251, J MOL BIOL
NAKASHIMA H, 1986, V99, P153, J BIOCHEM-TOKYO
NEEDLEMAN SB, 1970, V48, P443, J MOL BIOL
NISHIKAWA K, 1986, V871, P45, BIOCHIM BIOPHYS ACTA
PEARSON WR, 1988, V85, P2444, P NATL ACAD SCI USA
QIAN N, 1988, V202, P865, J MOL BIOL
RALPH WW, 1987, V3, P211, COMPUT APPL BIOSCI
ROBSON B, 1971, V58, P237, J MOL BIOL
ROOMAN MJ, 1988, V335, P45, NATURE
ROST B, 1993, V18, P120, TRENDS BIOCHEM SCI
SANDER C, 1991, V9, P56, PROTEINS
STOLORZ P, 1992, V225, P363, J MOL BIOL
SWEET RM, 1986, V25, P1565, BIOPOLYMERS
VISWANADHAN VN, 1991, V30, P1164, BIOCHEMISTRY-US
ZVELEBIL MJ, 1987, V195, P957, J MOL BIOL

20/5/27 (Item 24 from file: 34)
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci

02292738 Genuine Article#: KQ319 Number of References: 35
Title: FOUNDATION - A PROGRAM TO RETRIEVE ALL POSSIBLE STRUCTURES CONTAINING A USER-DEFINED MINIMUM NUMBER OF MATCHING QUERY **ELEMENTS FROM 3-DIMENSIONAL** DATABASES
Author(s): HO CMW; MARSHALL GR
Corporate Source: WASHINGTON UNIV,CTR MOLEC DESIGN/ST LOUIS//MO/63130; WASHINGTON UNIV,CTR MOLEC DESIGN/ST LOUIS//MO/63130
Journal: JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN, 1993, V7, N1 (FEB), P 3-22
ISSN: 0920-654X
Language: ENGLISH Document Type: ARTICLE
Geographic Location: USA
Subfile: SciSearch; CC LIFE--Current Contents, Life Sciences
Journal Subject Category: BIOCHEMISTRY & MOLECULAR BIOLOGY
Abstract: A program is described that **searches** three-dimensional, structural **databases** , given a user-defined **query** , in order to retrieve all structures that contain any combination of a user-specified minimum number of **matching** elements. Queries consist of three-dimensional coordinates of atoms and/or bonds. Numerous **query** constraints are described which allow the investigator to define the chemical nature of the desired structures as well as the environment within which these structures must reside. They include: (1) Bonded vs. isolated atom distinction; (2) Atom type designation; (3) Definition of subsets with occupancy specification (>, =, < X atoms); (4) RMS-fit; (5) Active site volume accessibility of atoms linking **query** elements, (6) Number, atom type, and cyclic structure constraints for atoms linking pharmacophoric elements; (7) Automatic error boundary adjustment - ad infinitum constraint.

To illustrate the capabilities of this program, queries based on the crystal structure of a thermolysin-inhibitor complex were tested **against** a subset of the Cambridge Crystallographic **Database** . Several compounds were returned which satisfied various aspects of the **query** , including fitting, within the active site. Combination of segments of compounds which satisfy **partial** queries should provide a method for generating unique compounds with affinity for sites of known three-dimensional structure.
Descriptors--Author Keywords: DRUG DESIGN ; **DATABASE SEARCHING** ; **CLIQUE** ALGORITHM ; FOUNDATION ; **MOLECULAR GRAPHICS**
Identifiers--KeyWords Plus: 3D CHEMICAL STRUCTURES; DRUG DESIGN; BINDING; FILES; INHIBITORS; ALGORITHM; DISPLAY; **SEARCH**; SITE
Research Fronts: 91-5517 001 (3-D COMPUTER VISION; CURVED SURFACES; RAY TRACING; INTERACTIVE PACKAGE)
Cited References:
ABOLA EE, 1985, ROLE DATA SCI PROGR
ALLEN FH, 1983, V16, P146, ACCOUNTS CHEM RES
ANDREWS PR, 1986, V4, P41, J MOL GRAPHICS
BARTLETT PA, 1989, P182, MOL RECOGNITION CHEM
BIERSTONE E, UNPUB CLIQUES GENERA
BOHM H, 1992, V6, P61, J COMPUT AID MOL DES
BONNER RE, 1964, V8, P22, IBM J RES DEV
BRINT AT, 1987, V5, P49, J MOL GRAPHICS
BRON C, 1973, V16, P575, COMMUN ACM
CORMEN TH, 1991, INTRO ALGORITHMS
CRANDELL CW, 1983, V23, P186, J CHEM INF COMP SCI
DESJARLAIS RL, 1988, V31, P722, J MED CHEM
DESJARLAIS RL, 1990, V87, P6644, P NATL ACAD SCI USA
FOLEY JD, 1982, FUNDAMENTALS INTERAC
GERHARDS L, 1981, V27, P349, COMPUTING
GIBBONS A, 1988, ALGORITHMIC GRAPH TH
GUND P, 1979, V14, P299, ANNU REP MED CHEM
GUND P, 1977, V5, P117, PROGR MOL SUBCELLULA
HO CMW, 1990, V4, P337, J COMPUT AID MOL DES

HOLDEN HM, 1987, V26, P8542, BIOCHEMISTRY-US
JAKES SE, 1986, V4, P12, J MOL GRAPHICS
JAKES SE, 1987, V5, P41, J MOL GRAPHICS
KUHL FS, 1984, V5, P24, J COMPUT CHEM
KUNTZ ID, 1982, V161, P269, J MOL BIOL
LESK AM, 1979, V22, P219, COMMUN ACM
LEWIS RA, 1992, V10, P66, J MOL GRAPHICS
MARTIN YC, 1992, V35, P2145, J MED CHEM
MARTIN YC, 1990, P213, REV COMPUTATIONAL CH
MOON J, 1991, V6, P314, PROTEIN-STRUCT FUNCT
NYBURG SC, 1974, V30, P251, ACTA CRYSTALLOGR B
SHERIDAN RP, 1987, V1, P243, J COMPUT AIDED DRUG
SHERIDAN RP, 1989, V86, P8165, P NATL ACAD SCI USA
SMELLIE AS, 1991, V31, P386, J CHEM INF COMP SCI
VANDRIE JH, 1989, V3, P225, J COMPUT AID MOL DES
VERLINDE CLM, 1992, V6, P131, J COMPUT AID MOL DES

```
Set     Items   Description
S1   15042460   MOLECULE? OR MOLECULAR OR PROTEIN? OR PEPTIDE? OR AMINO()A-
            CID? OR GENETIC? OR POLYPEPTIDE?
S2     288085   DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK? OR FILE?) -
            OR DB OR DBS OR DBMS OR RDB OR RDBM OR OODB?
S3    9339277   MATCH? OR COMPAR? OR QUERY OR QERIE? OR QUERYING OR SEARCH?
             OR LOCAT? OR FIND? OR SEEK?
S4    2334925   REPEAT? OR ITERAT? OR REITERAT? OR AGAIN?
S5    4851390   GRAPH? OR PARENT? OR INDEX OR INDICE? OR LIST? ?
S6    5616493   FRAGMENT? OR CLIQUE? OR PART OR PARTS OR PARTIAL OR SECTIO-
            N? OR STRING? OR SUBSTRING? OR MF OR MFS OR RESIDUE? OR CHAIN?
S7       1609   S2(2N)(POPULATE OR POPULATES OR POPULATING OR FILL OR FILLS
             OR FILLING OR BUILD OR BUILDS OR BUILDING OR CREATE OR CREAT-
            ES OR CREATING)
S8         46   S1(4N)S6 AND S7
S9         18   S1 AND S7 AND S3 AND S4 AND S5 AND S6
S10        61   S8 OR S9
S11        41   RD (unique items)
S12        20   S11 NOT PY>2000
S13        20   S12 NOT PD=20001117:20021117
S14        20   S13 NOT PD=20021117:20040501
File 305:Analytical Abstracts 1980-2004/Apr W2
         (c) 2004 Royal Soc Chemistry
File 399:CA SEARCH(R) 1967-2004/UD=14017
         (c) 2004 American Chemical Society
File  50:CAB Abstracts 1972-2004/Mar
         (c) 2004 CAB International
File  73:EMBASE 1974-2004/Apr W2
         (c) 2004 Elsevier Science B.V.
File  98:General Sci Abs/Full-Text 1984-2004/Apr
         (c) 2004 The HW Wilson Co.
File 143:Biol. & Agric. Index 1983-2004/Mar
         (c) 2004 The HW Wilson Co
File 154:MEDLINE(R) 1990-2004/Apr W2
         (c) format only 2004 The Dialog Corp.
File   5:Biosis Previews(R) 1969-2004/Apr W2
         (c) 2004 BIOSIS
File 285:BioBusiness(R) 1985-1998/Aug W1
         (c) 1998 BIOSIS
```

14/5/1 (Item 1 from file: 305)
DIALOG(R)File 305:Analytical Abstracts

342491 AA Accession No.: 64-31-A-10003 DOC. TYPE: Journal
A new approach to applications of the pattern recognition methods in analytical chemistry. IV. Automatic identification of structural fragments in organic compounds.
AUTHOR: Hippe, Z. S.. ; Kerste, A. ; Varmuza, K.
CORPORATE SOURCE: Univ. Information Technol. and Management, 35-225 Rzeszow, Poland
JOURNAL: Chem. Anal. (Warsaw), (Chemia Analityczna (Warsaw)), Volume: 46, Issue: 5, Page(s): 735-743
CODEN: CANWAJ ISSN: 0009-2223
PUBLICATION DATE: 2001 (1996200100) LANGUAGE: English
ABSTRACT: In this paper (part of a sequence devoted to automatic identification of organic substructures) a methodology of searching for optional classifiers for selected aromatic **fragments** embedded in organic **molecules** is briefly described. The developed methodology uses low-resolution mass spectra and employs computer program SCANKEE to **create** the **databases** for mass spectra and to search them to create spectrum-substructure correlation tables, and finally to convert automatically these tables into the rules database which enable effective concluding.
IDENTIFIERS: computer programs - SCANKEE, for pattern recognition based on structural fragments, in identn. of organic compounds, by MS ; mass spectrometry (MS) - in identn. of organic compounds, computer programs for
ANALYTE: organic compounds --identn. of, by MS, computer programs for
SECTION: A-20000 (General Analytical Chemistry)
SECTION CROSS-REFERENCE: C4 (Spectroscopy and Radiochemical Methods); D3 (Inorganic and Organic Analysis)

14/5/9 (Item 4 from file: 73)
DIALOG(R)File 73:EMBASE

05626265 EMBASE No: 1994040670
SOPM: A self-optimized method for protein secondary structure prediction
Geourjon C.; Deleage G.
Inst Biol et de Chimie des Proteines, UPR 412-CNRS, 7 Passage du Vercors,F-69367 Lyon cedex 07 France
Protein Engineering (PROTEIN ENG.) (United Kingdom) 1994, 7/2 (157-164)
CODEN: PRENE ISSN: 0269-2139
DOCUMENT TYPE: Journal; Article
LANGUAGE: ENGLISH SUMMARY LANGUAGE: ENGLISH

A new method called the self-optimized prediction method (SOPM) has been developed to improve the success rate in the prediction of the secondary structure of proteins. This new method has been checked against an updated release of the Kabsch and Sander database, 'DATABASE.DSSP', comprising 239 **protein chains** . The first step of the SOPM is to **build** sub- **databases** of protein sequences and their known secondary structures drawn from 'DATABASE.DSSP' by (i) making binary comparisons of all protein sequences and (ii) taking into account the prediction of structural classes of proteins. The second step is to submit each protein of the sub-database to a secondary structure prediction using a predictive algorithm based on sequence similarity. The third step is to iteratively determine the predictive parameters that optimize the prediction quality on the whole sub-database. The last step is to apply the final parameters to the query sequence. This new method correctly predicts 69% of amino acids for a three-state description of the secondary structure (alpha helix, beta sheet and coil) in the whole database (46 011 amino acids). The correlation coefficients are C(alpha) = 0.54, C(beta) = 0.50 and C(c) = 0.48. Root mean square deviations of 10% in the secondary structure content are obtained. Implications for the users are drawn so as to derive an accuracy at the amino acid level and provide the user with a guide for secondary structure prediction. The SOPM method is available by anonymous ftp to ibcp.fr.

MEDICAL DESCRIPTORS:
*protein secondary structure; *structure analysis
algorithm; amino acid sequence; article; comparative study; data base; priority journal; sequence analysis; statistical analysis; technique
SECTION HEADINGS:
029 Clinical and Experimental Biochemistry

14/5/17 (Item 2 from file: 154)
DIALOG(R)File 154:MEDLINE(R)

13591056 PMID: 9278278
Creation and characterization of a new, non-redundant fragment data bank.
Lessel U; Schomburg D
Gesellschaft fur Biotechnologische Forschung, Department of Molecular Structure Research, Braunschweig, Germany.
Protein engineering (ENGLAND) Jun 1997, 10 (6) p659-64, ISSN 0269-2139 Journal Code: 8801484
Document type: Journal Article
Languages: ENGLISH
Main Citation Owner: NLM
Record type: Completed
Subfile: INDEX MEDICUS
The success achieved for protein structure prediction of loop regions with insertions and deletions by knowledge-based methods depends on the quality of the underlying information, i.e. a fragment data bank as complete as possible is needed. However, the greater the number of proteins contributing to the data base the more redundant information is included, which leads to structurally similar proposals in loop predictions and to longer times for extracting fragments. So it is not only necessary to increase the number of proteins for **building** the loop **data base** but also to cluster the resulting fragments according to their structural similarities in order to remove redundancy. Here, a new, non-redundant fragment data bank is described, which is based on all proteins in the Brookhaven Protein Data Bank (release 7/95) with a resolution > or = 2.0 A and which can be updated easily by including new information from structures to be solved in the future. In the clustering process presented, the resulting clusters are optimized in several cycles until self-consistency. In this way all redundant information is removed without loosing any significantly different fragments. Finally the resulting fragment data bank is analysed with respect to its completeness.
Descriptors: Computational Biology--methods--MT; *Databases, Factual; * **Peptide Fragments** --analysis--AN; Algorithms; Amino Acid Sequence; Cluster Analysis; Protein Structure, Secondary; Protein Structure, Tertiary ; Sequence Homology, Amino Acid; Structure-Activity Relationship
CAS Registry No.: 0 (Peptide Fragments)
Record Date Created: 19971010
Record Date Completed: 19971010

14/5/20 (Item 1 from file: 5)
DIALOG(R)File 5:Biosis Previews(R)

0008794001 BIOSIS NO.: 199395096267
Chirbase: A molecular database for storage and retrieval of chromatographic chiral separations
AUTHOR: Roussel Christian; Piras Patrick
AUTHOR ADDRESS: ENSSPICAM, CNRS URA 1410, University Aix-Marseille III, 13397 Marseille Cedex 13, France**France
JOURNAL: Pure and Applied Chemistry 65 (2): p235-244 1993
ISSN: 0033-4545
DOCUMENT TYPE: Article
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: In order to meet the strong demand for storage and retrieval of chiral separations, we have developed Chirbase a **database build** on Chembase from Molecular Design Limited, a very powerful and well spread software. Chirbase allows the selection of the most promising conditions for a given chiral separation by searching and retrieving at the same time **molecular fragments** issued from the compound and from the stationary phase.

DESCRIPTORS:
MAJOR CONCEPTS: Biochemistry and Molecular Biophysics; Computer Applications--Computational Biology; Methods and Techniques
MISCELLANEOUS TERMS: ANALYTICAL METHOD
CONCEPT CODES:
00530 General biology - Information, documentation, retrieval and computer applications
10050 Biochemistry methods - General
10060 Biochemistry studies - General
10504 Biophysics - Methods and techniques

```
Set     Items   Description
S1     452321   MOLECULE? OR MOLECULAR OR PROTEIN? OR PEPTIDE? OR AMINO()A-
            CID? OR GENETIC? OR POLYPEPTIDE?
S2    1227560   DATABASE? OR DATABANK? OR DATA()(BASE? OR BANK? OR FILE?) -
            OR DB OR DBS OR DBMS OR RDB OR RDBM OR OODB?
S3    5904365   MATCH? OR COMPAR? OR QUERY OR QERIE? OR QUERYING OR SEARCH?
             OR LOCAT? OR FIND? OR SEEK?
S4    2338648   REPEAT? OR ITERAT? OR REITERAT? OR AGAIN?
S5    2824983   GRAPH? OR PARENT? OR INDEX OR INDICE? OR LIST? ?
S6    4759642   FRAGMENT? OR CLIQUE? OR PART OR PARTS OR PARTIAL OR SECTIO-
            N? OR STRING? OR SUBSTRING? OR MF OR MFS OR RESIDUE? OR CHAIN?
S7      30738   S2(2N)(POPULATE OR POPULATES OR POPULATING OR FILL OR FILLS
             OR FILLING OR BUILD OR BUILDS OR BUILDING OR CREATE OR CREAT-
            ES OR CREATING)
S8          4   S1(4N)S6(S)S7
S9          8   S1(S)S7(S)S3(S)S6
S10        36   S1(S)S6(S)S7
S11        43   S1(S)S2(S)S3(S)S4(S)S5(S)S6
S12       139   S1(4N)S7
S13       323   S6(4N)S7
S14         0   S11 AND (S12 OR S13)
S15         6   S1(10N)S2(10N)S3(10N)S4(10N)S5(10N)S6
S16        43   S8 OR S9 OR S10 OR S15
S17        35   RD (unique items)
S18        21   S17 NOT PY>2000
S19        21   S18 NOT PD=20001117:20021117
S20        21   S19 NOT PD=20021117:20040501
File 275:Gale Group Computer DB(TM) 1983-2004/Apr 20
         (c) 2004 The Gale Group
File 647:CMP  Computer Fulltext 1988-2004/Apr W2
         (c) 2004 CMP Media, LLC
File 674:Computer News Fulltext 1989-2004/Apr W2
         (c) 2004 IDG Communications
File 370:Science 1996-1999/Jul W3
         (c) 1999 AAAS
File 369:New Scientist 1994-2004/Apr W2
         (c) 2004 Reed Business Information Ltd.
File  16:Gale Group PROMT(R) 1990-2004/Apr 20
         (c) 2004 The Gale Group
File 160:Gale Group PROMT(R) 1972-1989
         (c) 1999 The Gale Group
File   9:Business & Industry(R) Jul/1994-2004/Apr 19
         (c) 2004  The Gale Group
File 129:PHIND(Archival) 1980-2004/Apr W2
         (c) 2004 PJB Publications, Ltd.
File 135:NewsRx Weekly Reports 1995-2004/Apr W2
         (c) 2004 NewsRx
File 429:Adis Newsletters(Archive) 1982-2004/Apr 20
         (c) 2004 ADI BV.
File 636:Gale Group Newsletter DB(TM) 1987-2004/Apr 20
         (c) 2004 The Gale Group
```

20/3,K/2 (Item 2 from file: 275)

02363085 SUPPLIER NUMBER: 58545234 (USE FORMAT 7 OR 9 FOR FULL TEXT)
Scientific and Technical Information: This Millennium and the Next. (News Briefs)
Lambert, Nancy
Searcher: The Magazine for Database Professionals, 8, 1, 24
Jan, 2000
ISSN: 1070-4795 LANGUAGE: English RECORD TYPE: Fulltext
WORD COUNT: 6904 LINE COUNT: 00559

... Business Information Services, the Procter & Gamble Company
At the end of the 20th century we **find** ourselves with a plethora of systems for **searching** the chemical structures in patents. The Derwent fragmentation code for non-polymeric structures has code terms applicable from 1963, 1970, 1972, and 1981. The time-ranged fragment coding is **searched** directly in the bibliographic World Patents **Index databases** (DWPI). There are two different chemical **fragmentation** codes for structures in the IFI CLAIMS US patents encoded between 1972 and the present. The IFI **fragments** must be **searched** for specific registered compounds in the CLAIMS Reference file and crossed over to the bibliographic UDB and CDB files, where the **fragmentation** code strategy is **searched again** for generic structures and infrequently encountered **molecules** . Chemical Abstracts Registry file has topological indexing of specific compounds, indeed, from patents since 1957...

...to the bibliographic CA and CAOLD files. Topological indexing of patents published since 1988 are **searched** directly in the companion MARPAT file. The Questel orbit **search** service offers topological **searching** with the Markush DARC system of the Merged Markush Service, which contains indexing of patents...

20/3,K/4 (Item 2 from file: 370)
DIALOG(R)File 370:Science

00500562 (USE 9 FOR FULLTEXT)
Mapping the Protein Universe
Holm, Liisa; Sander, Chris
The authors are in the European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton Hall, Cambridge CB10 1SD, UK.
Science Vol. 273 5275 pp. 595
Publication Date: 8-02-1996 (960802) Publication Year: 1996
Document Type: Journal ISSN: 0036-8075
Language: English
Section Heading: Articles
Word Count: 6817

(THIS IS THE FULLTEXT)

...Text: misleading when subtle irregularities in the coordinates lead to spurious differences in these vectors for **proteins** that are actually similar in shape. The algorithm works by storing, in a way convenient for geometrical lookup, a **list** of spatial relations between such vectors taken from **database** proteins (B8) . Here, lookup (or "hashing") is conceptually similar to looking up names in a telephone book. The lookup procedure **matches** the vector relations taken from the **query** protein with those in the stored **list** and proceeds to sample a limited set of spatial superimpositions whenever enough **matches** are found between the **query protein** and a **database protein** . Finally, a dynamic programming step refines these superimpositions and generates detailed **residue** -level alignments. The **search** of one structure **against** the structure **database** of several thousand structures typically takes only about 5 min on a computer workstation. Other...

...achieve similar speed (B7) . In this way, a large portion (about 90%) of all significant **protein - protein** shape similarities can be found (Fig. 3A...

20/3,K/7 (Item 3 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)

07469537 Supplier Number: 62767418 (USE FORMAT 7 FOR FULLTEXT)
Assembly Required.(genetic research done by Cell Map)
Moukheiber, Zina
Forbes, p132
July 3, 2000
Language: English Record Type: Fulltext
Document Type: Magazine/Journal; General Trade
Word Count: 1244

... measured changes in thousands of proteins in healthy and diseased spinal fluids before narrowing the **list** down to 10 to 20 **proteins** strongly correlated with memory loss. Arobotic arm carved out the **proteins** of interest, grabbing each sample and breaking it into **fragments** . It loaded the **fragments** into a mass-spec machine for sequencing.

OGS **matched** those **proteins** **against** Incyte's LifeSeq gene **database** and three public **databases** . It found **proteins** never seen before in Alzheimer's. Altogether, it took OGSa year to complete the work ...

20/3,K/15 (Item 11 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)

04349668 Supplier Number: 46379595 (USE FORMAT 7 FOR FULLTEXT)
INCYTE LAUNCHES VERSION 4.0 OF THE LIFESEQ DATABASE New Release Includes Public-Domain DNA Sequence Data
News Release, pN/A
May 13, 1996
Language: English Record Type: Fulltext
Document Type: Magazine/Journal; Trade
Word Count: 1224

(USE FORMAT 7 FOR FULLTEXT)
TEXT:
...powerful bioanalysis tools. The database contains information generated by analyzing more than 1 million gene **fragments** , representing approximately 100,000 distinct human genes. This drag-discovery tool will be used by...

...4.0 incorporates extensive cross-referencing between Incyte sequences and GenBank, the repository of public **genetic** information sponsored by the National Center for Biotechnology Information (NCBI). The resulting LIFESEQ annotations are...
...DNA Sequence database with the Gene Expression database. This enables scientists to manipulate DNA or **protein** sequence alignments and integrate them with the gene-expression profiles of different tissues and cell...

...Incyte's goal is to make the LIFESEQ database the product of choice for scientists **seeking** to analyze and manage both proprietary and public genomic data sets. Toward this end, Incyte...

...Marketing. "Scientists can use LIFESEQ to perform the electronic equivalents of biological experiments, such as **comparing** the gene-expression profiles of 'normal' and 'diseased' tissues. Each of these electronic analyses takes just seconds in the computer, **compared** with weeks of work in a traditional laboratory." What is LIFESEQ? The LIFESEQ database is...

...largest and most powerful collections of human genomic data. It provides a picture of cellular **genetics** at a level of detail never before possible, helping researchers determine which genes, both known...

...the way pharmaceutical companies conduct research, develop drugs, and even diagnose and treat diseases. In **building** the LIFESEQ **database** , Incyte harnesses the power of high-throughput sequencing to decipher the structure of DNA (deoxyribonucleic acid), the **molecule** that makes up our chromosomes and determines heredity. It then uses sophisticated bioanalysis software to...

...access to robust sequence-analysis tools such as BLAST, which allows researchers to sort and **search** the data in their quest for promising new drag targets. The Gene Expression database contains...

...of "point-and-click" biology. For example, with just a few mouse clicks, scientists can **compare** the genes functioning in healthy prostate tissue with those active in prostate cancer. In addition...

...from Incyte's consultation with our scientists to enhance the product's integrated approach to **genetic** database mining for target identification, confirmation, and validation?' Other Database Modules To complement and expand...

...Incyte is developing new generations of database modules. The Gene Mapping module identifies the chromosomal **locations** for selected gene sequences and promises to be a valuable resource in the hunt for...Its LIFESEQ and Gehe Mapping databases integrate bioinformatics software with both proprietary and publicly available **genetic** information to create an

information-based tool used by pharmaceutical companies in drug discovery and...